BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2017-45

Identification of Random Resource Shares in Collective Households Without Preference Similarity Restrictions

by Geoffrey R. Dunbar, Arthur Lewbel and Krishna Pendakur

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2017-45

October 2017

Identification of Random Resource Shares in Collective Households Without Preference Similarity Restrictions

by

Geoffrey R. Dunbar,[1] Arthur Lewbel[2] and Krishna Pendakur[3]

[1] International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gdunbar@bankofcanada.ca

[2] Department of Economics, Boston College
Chestnut Hill, Massachusetts, United States 02467
lewbel@bc.edu

[3] Department of Economics, Simon Fraser University
Burnaby, British Columbia, Canada V5A 1S6
pendakur@sfu.ca

ISSN 1701-9397

© 2017 Bank of Canada

Abstract

Resource shares, defined as the fraction of total household spending going to each person in a household, are important for assessing individual material well-being, inequality and poverty. They are difficult to identify because consumption is measured typically at the household level, and many goods are jointly consumed, so that individual-level consumption in multi-person households is not directly observed. We consider random resource shares, which vary across observationally identical households. We provide theorems that identify the distribution of random resource shares across households, including children's shares. We also provide a new method of identifying the level of fixed or random resource shares that does not require previously needed preference similarity restrictions or marriage market assumptions. Our results can be applied to data with or without price variation. We apply our results to households in Malawi, estimating the distributions of child and female poverty across households.

Bank topics: Domestic demand and components; Econometric and statistical methods
JEL codes: D13, D11, D12, C31, I32

Résumé

La part de ressources, au sens de fraction de la dépense totale d'un ménage qui se rapporte à chaque membre du foyer, joue un rôle important dans l'évaluation du bien-être matériel des personnes, de leur inégalité de situation et de leur niveau de pauvreté. Il est difficile d'estimer cette part, car la consommation est généralement mesurée au niveau du ménage. De plus, comme les personnes d'un même foyer consomment ensemble bon nombre de biens, la consommation de chacun dans des ménages de plus d'un individu n'est pas directement observable. Nous examinons des parts de ressources qui varient aléatoirement dans des ménages au profil identique. Nous présentons des théorèmes qui permettent de dégager la répartition des parts aléatoires de ressources parmi les ménages recensés, notamment celles des enfants. Nous exposons également une nouvelle méthode de calcul des parts fixes ou aléatoires de ressources, qui ne demande pas comme auparavant l'imposition de restrictions sur la similarité des préférences ni la formulation d'hypothèses sur le marché matrimonial. Nos résultats sont applicables aux données assorties ou non d'une variation de prix. Nous les appliquons à des ménages du Malawi pour évaluer la répartition des taux de pauvreté des femmes et des enfants.

Sujets : Demande intérieure et composantes; Méthodes économétriques et statistiques
Codes JEL : D13, D11, D12, C31, I32

Non-Technical Summary

Expenditure surveys generally collect consumption data at the level of households. Standard poverty and welfare measurements based on such data are also typically calculated at the household level. But well-being and utility apply to individuals, not households. When household resources are distributed unequally across household members, official household level measures of income or consumption can seriously underestimate the prevalence of poverty and inequality in a country.

The *resource share* of a given household member is defined as the fraction of total household consumption expenditures that are spent on goods and services that the given household member consumes. Resource shares are in general difficult to identify, particularly children's resource shares, because consumption is measured typically at the household level, and many goods are jointly consumed and/or shareable.

In this paper we provide two new theorems on the identification of resource shares within households, and provide associated estimators. Our first theorem allows for unobserved heterogeneity in resource allocations across households. By treating resource shares as random variables, our first theorem shows identification of the distribution (around an unknown level) of resource shares across households. Our second theorem shows identification of the level of resource shares. Unlike existing related theorems, this second theorem does not require assumptions regarding similarity of tastes across households of different composition or across people within a household. This second theorem instead uses a new way to exploit variation in so-called *distribution factors*, that is, variables that affect the allocation of resources within households but do not affect the tastes of individuals. Taken together, our theorems completely identify the joint distribution of resource shares across households.

We apply our results to household survey data from Malawi. We find that allowing for unobserved heterogeneity in resource allocations across households matters for welfare analysis in two important ways. First, we find that unobserved heterogeneity accounts for about the same amount of variation in resource shares as does observed heterogeneity. Second, we find that measured poverty rates are lower when allowing for random resource shares rather than constraining them to be fixed.

In addition to identifying the level and distribution of resource shares, we also use our results to test some of the behavioural restrictions that previous work required for identification. We find evidence supporting the similarity across people (SAP) restriction that was proposed by Dunbar, Lewbel and Pendakur (2013) for identification.

1 Introduction

Expenditure surveys generally collect consumption data at the level of households. Standard poverty and welfare measurements based on such data are also typically calculated at the household level. But well-being and utility apply to individuals, not households. When household resources are distributed unequally across household members, official household-level measures of income or consumption can seriously underestimate the prevalence of poverty and inequality in a country. For example, using the types of methods that we will extend here, Dunbar, Lewbel and Pendakur (2013, hereafter denoted DLP), find poverty rates for children in Malawi are much higher than those of men. Another example is Calvi (2017), who finds that the poverty rate among older married women in India increases with age, primarily because their share of household resources declines with age.

The *resource share* of a given household member is defined as the fraction of total household consumption expenditures that are spent on goods and services that the given household member consumes. Resource shares are in general difficult to identify, particularly children's resource shares, because consumption is measured typically at the household level, and many goods are jointly consumed and/or shareable.

In this paper we provide two new theorems on the identification of resource shares within households, and provide associated estimators. Our first theorem allows for unobserved heterogeneity in resource allocations across households. By treating resource shares as random variables, our first theorem shows identification of the distribution (around an unknown level) of resource shares across households. Our second theorem shows identification of the level of resource shares. Unlike existing related theorems, this second theorem does not require assumptions regarding similarity of tastes across households of different composition or across people within a household. This second theorem instead uses a new way to exploit variation in so-called *distribution factors*, that is, variables that affect the allocation of resources within households but do not affect the tastes of individuals. Taken together, our theorems completely identify the joint distribution of resource shares across households.

We apply our results to household survey data from Malawi. We find that allowing for unobserved heterogeneity in resource allocations across households matters for welfare analysis in two important ways. First, we find that unobserved heterogeneity accounts for about the same amount of variation in resource shares as does observed heterogeneity. Second, we find that measured poverty rates are lower when allowing for random resource shares rather than constraining them to be fixed.

In addition to identifying the level and distribution of resource shares, we also use our results to test some of the behavioural restrictions that previous work required for identifi-

cation. We find evidence supporting the similarity across people (SAP) restriction that was proposed by DLP for identification.

We start with the Pareto efficient type of collective household model of Becker (1965, 1981) and Chiappori (1988, 1992). The earlier literature on these models, including Browning, Bourguignon, Chiappori, and Lechene (1994), Browning and Chiappori (1998), Vermeulen (2002), and Chiappori and Ekeland (2009), showed that, even if one knew the entire demand vector-function of a household (that is, how much the household buys of every good and service as a function of prices, income and other observed covariates), one still cannot identify the level of each household member's resource. However, this earlier work also shows that one can generally identify how these resource shares would change in response to a change in observed covariates such as distribution factors. Other papers that make use of this result include Bourguignon and Chiappori (1992), Chiappori, Fortin and Lacroix (2002), and Blundell, Chiappori and Meghir (2005). Most of this earlier work also constrains goods to be either purely private or purely public within a household, whereas we work with a more general model based on Browning, Chiappori and Lewbel (2013) and DLP, that allows some or all goods to be partly or fully shared.

Our first theorem extends this earlier literature by showing that, though the level of resource shares is not identified without additional assumptions, one can identify the distribution of random variation in these shares (around the unknown levels) across households. Two prior results exist on identification of the distribution of random resource shares, both of which are much more restrictive than ours. One, by Matzkin and Perez-Estrada (2011), is based on semiparametric restrictions of functional forms and focuses on variation in a single good. The other, by Chiappori and Kim (2017), is closer to ours but unlike our Theorem 1, requires observed price variation, deals only with two-member households, and requires every good to be either purely private or purely public.

As noted above, from just observing household demand functions, one cannot identify the absolute levels of resource shares; only variations in resource shares around those levels are identified. Some interesting policy questions can be addressed without identifying levels, such as identifying whether a policy that changes a distribution factor, like women's access to contraception, actually increases women's share of resources within the household. However, other fundamental policy questions, such as identifying the prevalence of women's or children's poverty, requires identifying resource share levels. Our Theorem 2 provides minimal additional assumptions needed to identify the levels of resource shares.

A number of previous approaches exist to address the fundamental nonidentification of resource share levels just from household demand data. One direct approach, taken, e.g., by Menon, Pendakur and Perali (2012) and Cherchye, De Rock and Vermeulen (2012), is

to collect as much detailed data as possible on the separate consumption of each household member, rather than just observing household-level consumption. To the extent that such data can be collected, resource shares may be observed directly, by taking each individual's observed consumption as a fraction of the total. However, this method requires detailed and difficult data collection, and is likely to suffer from considerable measurement errors, particularly in the allocation of public and shared goods to individual household members.

A second approach is taken by Cherchye, De Rock and Vermeulen (2011). While the levels of resource shares cannot be identified without additional assumptions, these authors show that it is possible to obtain bounds on resource shares, using revealed preference inequalities. Cherchye, De Rock, Lewbel, and Vermeulen (2015) considerably tighten these bounds by combining Slutsky-symmetry-type restrictions with revealed preference inequalities. Bounds can be further tightened, sometimes leading to point identification of shares, by further combining revealed preference inequalities with assumed restrictions regarding marriage markets. See, e.g., Cherchye, De Rock, Demuynck, and Vermeulen (2014).

A third method is to completely identify the level of resource shares from household-level data by imposing additional restrictions either on preferences or on the household's allocation process, or both. Papers that use these methods include Lewbel (2003), Lewbel and Pendakur (2008), Couprie, Peluso and Tranmoy (2010), Bargain and Donni (2009, 2012), Lise and Seitz (2011), Browning, Chiappori and Lewbel (2013), and DLP. These methods all employ either strong restrictions on the functional forms of preferences, or assumptions regarding similarity of preferences across households of different composition or similarity of tastes across people within a household. For example, Lise and Seitz (2011) impose a particular functional form on utility functions, Browning, Chiappori and Lewbel (2013) assume individuals' indifference curves over goods do not change when they marry, and DLP assume tastes for an assignable good like clothing are similar either across people within a household, or across people of the same type across households. Most of these papers that impose preference restrictions on individuals do not require or exploit distribution factors.

Both Lewbel and Pendakur (2008) and DLP make an additional assumption that resource shares do not depend on the total expenditures of the household (though they can still depend on the total income or wealth of the household). This assumption allows them to use data without price variation, i.e., to work with Engel curves instead of Marshallian demand functions. This considerably reduces data requirements (relative to all other papers that identify resource share levels) by: (i) greatly simplifying the demand functions that need to be estimated, (ii) eliminating the need to collect price data, and, (iii) allowing estimation using data from a single cross-section survey. Evidence supporting this assumption empirically can be found in intra-household data collection papers like Menon, Pendakur and Perali (2012)

and in the bounds estimates of Cherchye, De Rock, Lewbel and Vermeulen (2015).

Our second theorem provides a new set of minimal conditions that suffice for identification of the level of resource shares. This theorem shows that if one maintains just the assumption that resource shares do not depend on total expenditures, and observes some assignable goods and some distribution factors, then that alone can be enough to identify the level of resource shares. No similarity restrictions on tastes like those discussed above are needed. Moreover, like Lewbel and Pendakur (2008) and DLP, these conditions permit identification without requiring any price variation. Because this new identification result does not impose these preference similarity restrictions, it allows us to empirically test those constraints.

After establishing identification, we show how these models can be estimated and tested. A general semiparametric method of estimating individuals' demand functions within a household, along with the distribution of resource shares, is given in the appendix. In the main text we provide a simpler estimator that combines a commonly used functional form for collective household model demands with nonparametric estimation of the random variation in resource shares.

The next sections set up the model and provide our identification theorems. We then describe our proposed estimators. This is followed by empirical results from applying our model to a survey of households in Malawi. We use the results to test the types of preference restrictions that earlier identification results required. We also report estimates of the distribution of child poverty (and female poverty) across households, which account for random variation in resource shares.

2 Resource Shares

A key component of collective household models, going back to the earliest frameworks of Becker (1965, 1981) and Chiappori (1988, 1992), is resource shares. Resource shares are defined as the fraction of a household's total resources (spent on consumption goods) that are allocated to each household member. Resource shares, which are closely related to Pareto weights, are often interpreted as measures of the bargaining power of each household member; however, they are also determined by altruism, particularly the shares claimed by children. Our model starts with the Pareto efficient collective household model of Browning, Chiappori and Lewbel (2013, hereafter denoted BCL).

2.1 BCL and DLP

Unlike earlier collective household models, BCL's model does not require goods to be purely public or purely private, but instead permits goods to be partly shared, using what is called a consumption technology function. Like earlier results in this literature, BCL show that as a result of the Pareto efficiency of the household's resource allocation process, maximizing the household's objective function is observationally equivalent to a decentralized allocation. In this allocation, each household member demands a vector of consumption quantities given their preferences and a personal budget constraint, and the household purchases the sum of these demanded quantities (adjusted to account for the extent to which goods are shared or consumed jointly).

Given this decentralization, we can conceptualize the household's behaviour as creating budget constraints for its members. These budget constraints, combined with the preferences of individual household members, are sufficient to conduct ordinal welfare analysis. Each person's budget constraint is characterized by a shadow budget and a shadow price vector. These are "shadow" budgets and prices because they govern each agent's consumption demands but are not observed and do not equal the observed household budget or market prices.

Shadow budgets must add up to the full household budget. Each person's share of the household budget is called their *resource share*. Resource shares may be unequal across household members, due to individual variation in bargaining power and altruism within the household. BCL show that resource shares have a one-to-one correspondence with Pareto weights on individual utilities in the overall household maximization problem. Resource shares may depend on *distribution factors*, defined as variables that affect bargaining power or claims on resources within the household, but do not affect preferences for goods and services.

In the BCL model, shadow prices for goods must be the same for all household members. (If they were not the same, then there would be gains from trade across household members, a violation of the assumption of efficiency.) Shadow prices are determined by the consumption technology of the household, and are less than or equal to market prices. The more a good is shared, the lower is its shadow price. For goods that are not shared, the shadow price equals the market price. In their empirical application, BCL impose the restriction that the shadow price vector is a linear transformation of market prices. We do the same here. This ensures that a linear household budget constraint given market prices implies a linear shadow budget constraint for each household member.

BCL fully identify both the consumption technology function and the resource shares, so that they can completely characterize the shadow budget constraints faced by each house-

hold member. They accomplish this by imposing strong restrictions on the similarity of preferences between individuals living alone vs living together, requiring data on both the demands of singles and of couples, and by requiring substantial observed relative price variation. More recent work focuses on relaxing these assumptions and data requirements, and as a result only on identifying some particular features of interest, such as resource shares. Our paper falls in this category.

Lewbel and Pendakur (2008) and DLP modified BCL to permit identification of resource shares from data that do not contain price variation (Engel-curve-type data), by placing restrictions on how prices and the consumption technology function interact, and by imposing the constraint that resource shares not vary with total expenditures. Both provide theoretical and empirical evidence supporting this identifying assumption. DLP also substantially relaxed the BCL restriction limiting differences in preferences, and further relaxed data and estimation requirements by only needing to observe and estimate household Engel curves on one private assignable good for each household member. Note that private goods are defined as goods that are not shared, and an assignable good is one where we can observe which household member consumes the good.

Unlike BCL, DLP permits identification of resource shares of both adults and children, where children are treated as having their own utility functions and welfare. This is in contrast to most of the empirical collective household literature, where expenditures on children are modelled just as public goods in the adults' utility functions. The identification of children's resource shares is necessary to answer questions regarding the welfare of children in the household, separate from the welfare of the parents.

The present paper, like DLP, is a model with both adults' and children's resource shares, in a data environment based on observable assignable goods that does not require price variation. As discussed in the introduction, the present paper generalizes DLP in two important ways. One is the identification of resource shares without imposing the BCL or DLP restrictions on similarity of preferences across households of different composition, or across people within a household. The other is by allowing resource shares to vary randomly across households, equivalent to allowing for the existence of unobserved distribution factors, or unobserved heterogeneity in bargaining power or altruism.

2.2 Children and Random Resource Shares

Let d denote a vector of distribution factors. Distribution factors are important variables in the collective household literature for three reasons. First, distribution factors are closely related to individuals' relative bargaining power within households, and so are important

components of marriage market models, and models of women's empowerment (particularly in poor and developing countries). As a result, distribution factors arise in other literatures associated with household formation, stability and function. Second, some types of distribution factors are policy variables, providing governments with the ability to affect the within-household distribution of resources. Examples can include inheritance and divorce laws, access to credit or banking for women, and local availability of education or medical care. Third, in efficient collective households the response of resource shares to a change in distribution factors can generally be identified just from household-level demand behaviour, given many observations of the same or fully comparable household(s) in different price and income regimes, without additional assumptions.

Our model allows for random variation in resource shares across households. Equivalently, we allow for the presence of unobserved distribution factors, which we will denote v. This reflects the fact that the relative bargaining power of individual household members, or the choice of social welfare function for the household, will in general vary across households for unobserved reasons. So, e.g., two households that otherwise appear identical could have different values of resource shares, due to unobserved differences in household members' personality traits, beauty, intelligence, altruism, etc. Both d and v are defined to be variables that affect resource shares, but do not affect preferences.

Let the household be comprised of J individuals indexed by j. Our typical case will be $J = 3$, consisting of a man, a woman and children.[1] Let x^* be the total expenditures of the household, i.e., the household's total budget.

Let z denote a vector of observable attributes of households and their members like age, education and number of children. Household attributes z may affect the preferences of each household member and may also affect the household's bargaining process or social welfare function, and as a result may directly affect resource shares.

Let scalars $y_1,...,y_J$ denote quantities of private assignable goods. Specifically, each y_j is the quantity of some good j that is only consumed by household member j. What makes each y_j assignable is that each appears in just one (known) person's utility function, and what makes each y_j private is that it is not shared or consumed jointly. So, e.g., in a household with a woman, a man and a child, the private assignable goods could be women's clothes, men's clothes and children's clothes.

Let p denote the vector of prices of all the goods and services that the household buys, sorted so that the prices of assignable goods are the first J elements of p. Thus, p includes

[1] Due to data limitations in our empirical analysis, we treat all of the children as having a single utility function, or equivalently, a unitary joint social welfare function. Other datasets could potentially allow researchers to relax this restriction, by observing some component of consumption separately for each child.

the prices $p_1,...,p_J$ of the private assignable goods, as well as the prices of all other goods the household buys. Let $\eta_j(p, x^*, z, d, v)$ denote the resource share of member j, that is, η_j is the fraction of total household resources x^* that are allocated to member j.

Adapting BCL and DLP, in the Appendix we show that the household demand functions of private assignable goods are given by

$$q_j = h_j(\eta_j(p, x^*, z, d, v)x^*, p, z) \quad \text{for } j = 1, ..., J. \tag{1}$$

The interpretation of this equation is that the resources allocated to member j are $\eta_j x^*$ (the share η_j of total household budget x^*) and the function h_j is that member's Marshallian demand function for this good.[2] Since this good is private and assignable, the household's demand for the good equals just that member's own demand for the good. It is important to note that *only* private assignable goods have the simple form given by equation (1). The demand functions for other goods are much more complicated, as in BCL.

We will assume that resource shares η_j are independent of the household budget x^*. Lise and Seitz (2011), Lewbel and Pendakur (2008), Bargain and Donni (2009, 2012) and DLP all use this restriction in their identification results, and supply some theoretical arguments for it. Cherchye, De Rock, Lewbel and Vermeulen (2015) and Menon, Pendakur and Perali (2012) provide empirical support for this restriction using American and Italian data, respectively. Also reducing the restrictiveness of this assumption is that resource shares are only assumed to be independent of x^* after conditioning on covariates such as z and d, both of which could include variables closely related to x^*, such as wealth, income (which equals x^* plus savings), education, wages, etc.

Our model allows for a certain type of measurement error in the household budget. Let x denote the observed household budget, which is related to the true household budget x^* by

$$x^* = cx,$$

where c is measurement error, which we assume has mean one. We assume this error is of the Berkson (1950) type, meaning that it is distributed independently of x. We assume measurement error takes a multiplicative form because larger levels of expenditures are likelier to be associated with larger errors. Our main reason for assuming Berkson errors is pragmatic; with our particular model and identifying assumptions, we can allow for this type of error essentially for free, i.e., with no additional assumptions and no added complexity in

[2]Note that this h_j function may depend on z not just because tastes may vary with z, but also because households can vary in the extent to which they share goods other than the private assignable goods, and so the h_j functions include accounting for the resulting variation in shadow prices relative to market prices p.

estimation. In fact, in our empirical estimates reported later, we find little evidence of these errors. It may be possible to also allow for ordinary (classical) measurement error using a control function approach, though we do not pursue that possibility here.

The assumption that resource shares are independent of the household budget means we may drop x^* from the η_j function, and our total expenditure measurement error model means we can replace x^* with cx in equation (1). The result is

$$w_j = h_j(\eta_j(p, z, d, v), cx, p, z) \quad \text{for } j = 1, ..., J. \tag{2}$$

We assume that we observe data $\{w_1, ..., w_J, p, z, d, x\}$ for a sample of households. However, if all households in the sample face the same price vector, as in a cross section of Engel curve data, then it is unnecessary to observe p, and all of our results will go through just conditioning on that period's prices.

The random unobservables in our model consist of the measurement error c and the unobserved distribution factors v. Since the resource shares $\eta_1, ..., \eta_J$ depend on the unobserved random vector v, these resource shares are themselves random variables. Our goal is to nonparametrically identify the conditional distribution of $\eta_1, ..., \eta_J$, conditioning on z, d and, if price variation is observed, p. The identification is nonparametric in that the theory assumes the functional forms of $h_1, ..., h_J$ are unknown, and functional forms of how $\eta_1, ..., \eta_J$ depend on p, z, d and v are also unknown.

Our identification argument proceeds in two steps. First we show the distribution of $\eta_1, ..., \eta_J$ is identified around unknown conditional mean functions. The distribution of the measurement error c is also nonparametrically identified. Then we provide sufficient conditions for identifying these conditional mean functions, without imposing the DLP- or BCL-type preference similarity restrictions, by exploiting variation in d. The advantage of separating these two results is that identification of the conditional mean functions requires more assumptions. Also, behavioural restrictions like those in DLP or BCL could be used instead of, or in addition to, the method we provide here for identifying the conditional means of the resource shares.

In an appendix we describe a general estimator based on these identification results. In the main text we provide an estimator that is easier to implement, but is based on a particular popular functional form for the demand functions h_j. Both estimators are nonparametric regarding the the distribution of resource shares $\eta_1, ..., \eta_J$ and of the measurement error c.

3 Nonparametric Identification of Resource Shares and Their Distribution

In this section we first give some intuition for how we use observed distribution factors to identify the levels of deterministic resource shares, providing an alternative to the restrictions on preferences that were required by BCL, DLP and others to identify levels. We then allow resource shares to vary randomly across households (corresponding to the presence of unobserved distribution factors) and give some intuition for how we can identify the joint distribution of these resource shares. This is then followed by our formal list of assumptions and identification theorems.

3.1 Identification Intuition With Fixed Resource Shares

To simplify notation in this subsection, assume we have a sample with no variation in p, z, or d. Assume also for now that there is no measurement error, so $e = 1$, and no unobserved distribution factors v (and therefore no random variation in resource shares). Equation (2) then simplifies to

$$y_j = G_j(x) = h_j(\eta_j x) \quad \text{for} \quad j = 1, ..., J, \tag{3}$$

where $G_j(x)$ is an observable household demand function for person j's private assignable good. The goal is then identification of η_j, which is now just a constant for each household member j.

3.2 BCL and DLP Identification

In this greatly simplified version of the model, the BCL identification would correspond to identifying the h_j functions by looking at singles (people living alone, for whom $\eta_j = 1$). Then, with h_j known, one could solve equation (3) for η_j. This requires assuming that individuals' tastes for goods, and therefore the h_j functions (apart from the consumption technology) do not change when people form collective households (or, in particular, when they marry).

Note we are just providing intuition here; the actual identification in BCL is more difficult because (among other restrictions) they do not require the existence of private assignable goods, they do not impose the constraint that η_j not depend on x, and they explicitly account for the consumption technology (i.e., sharing) of other goods. Other papers like Lewbel and Pendakur (2008) and Bargain and Donni (2012) similarly make use of singles to identify h_j and thereby recover η_j.

DLP employ (two) different identification strategies that do not depend on singles. One of DLP's approaches is to instead suppose that h_j is similar across people (SAP) within the household. An extreme case would be if the h_j function was the same across all the $j = 1, ...J$ people in the household, so we'd have $y_j = h(\eta_j x)$ instead of $y_j = h_j(\eta_j x)$ for some function h. Here, there are $J-1$ unknown scalars η_j and one unknown function h to identify from J observed functions G_j, so we meet the order condition for identification.

In particular, variation in x allows us to identify η_j. To illustrate, suppose h was linear so $h(\eta_j x) = b\eta_j x + a$ for unknown coefficients b and a. Then the observable demand functions (for these private assignable goods) are $G_j(x) = c_j x + a$ where $c_j = b\eta_j$. The coefficient of x identifies c_j. Using the fact that resource shares η_j sum to one, we can identify b by $\sum_{j=1}^{J} c_j = \sum_{j=1}^{J} b\eta_j = b$. Then, once we know b and each c_j, we identify each η_j by $\eta_j = c_j/b$. Essentially, we identify the J coefficients $c_1, ...c_J$, and from those we obtain the J parameters $b, \eta_1, ...\eta_{J-1}$. The last parameter η_J is obtained from the constraint that shares sum to one.

For nonlinear h, we can do a similar construction with derivatives (denoted by apostrophes). For example, if $G_j(x) = h(\eta_j x)$ then $G_j'(x) = \eta_j h'(\eta_j x)$. So $G_j'(0) = \eta_j h'(0)$ and we can treat treat $h'(0)$ like b above. Again we are just providing intuition; the actual identification in DLP is more complicated than this illustration. For example, DLP allows for h_j functions that are not identical across people but merely similar (i.e., only having some coefficients in common). DLP also consider cases where, instead of being similar across household members, demands are similar across types (SAT), e.g., women having similar demands regardless of whether the household has one, two or three children.

3.3 Identification Using Observed Distribution Factors

Both BCL and DLP (and other related results in the literature) depend heavily on some preference restrictions, like tastes not changing when two individuals form a couple, or similarity of tastes across household members. But, unlike the earlier literature that identified only changes but not levels in resource shares, the newer identification results like those in BCL and DLP do not require distribution factors. Our new method for identifying levels of resource shares, given as Theorem 1 in the next section, does not impose the types of preference restrictions needed by results like BCL and DLP. Instead our new method exploits observable distribution factors while maintaining the assumption that resource shares do not depend on x.

For intuition on how Theorem 1 will work, continue to assume a sample with no variation in p or z, no measurement error and no unobserved distribution factors v, but now assume we see some variation in distribution factors d, so the assignable good demand functions are

now

$$y_j \quad G_j(d,x) = h_j(\eta_{dj}x) \quad \text{for } j = 1,...,J, \tag{4}$$

where $G_j(d,x)$ is an observable demand function. The goal is identification of η_{dj}, which now is just a different constant for each value that d and j can take on.

To keep things simple, suppose d is a discrete scalar that just takes the values $1, 2, ..., D$ for some integer D. Again, just for intuition, consider the case where these h_j demand functions are linear, but we no longer impose the DLP SAP restriction, so $h_j(\eta_{dj}x) = b_j\eta_{dj}x + a_j$ for unknown coefficients b_j and a_j. Then the observable household demand functions (for these private assignable goods) are $G_j(d,x) = c_{dj}x + a_j$ where $c_{dj} = b_j\eta_{dj}$. For each value d can take on, the coefficient of x in the j'th equation identifies the coefficient c_{dj}. We can identify DJ coefficients, i.e., c_{dj} for each value of d and j. The unknown parameters are now $b_1,...b_J$, and, for each value of d, $\eta_{d1},...\eta_{dJ-1}$ (again using that $\eta_{dJ} = 1 - \sum_{j=1}^{J-1}\eta_{dj}$). This gives a total of $J + (J-1)D$ unknowns. So if $D \geq J$, meaning that we have at least as many different values of the distribution factors as we have household members, then we have enough equations to identify all the resource shares η_{dj}.

The above shows we have at least as many equations as unknowns. This is a necessary but not sufficient condition for identification, so the identification proof also entails showing that a corresponding rank condition holds. For example, consider the case where $D = J = 2$. Then the equations $c_{dj} = b_j\eta_{dj}$ and $\eta_{d2} = 1 - \eta_{d1}$ together imply $c_{11}/c_{21} = \eta_{11}/\eta_{21}$ and $c_{12}/c_{22} = (1-\eta_{11})/(1-\eta_{21})$, which are two equations that, given some mild inequalities, we can readily solve for the two unknowns η_{11} and η_{21}.

Our actual identification theorem extends this idea to nonlinear demand functions by doing a similar analysis based on derivatives of $G_j(d,x)$ with respect to x in place of linear coefficients of x. Finally, in all of these results, everything we call constants above are, in the real models, functions of z, and also of p if the data contain price variation.

The main result here is that, instead of requiring preference similarity restrictions, all we need is an observable distribution factor vector d that can take on as many values as we have family member types. For example, if $J = 3$ (man, woman and children), then two binary distribution factors are enough.

3.4 Identification Intuition For Random Resource Shares

Now consider identification in the presence of unobserved distribution factors v, or equivalently, identification of the distribution of random resource shares. Again simplifying notation as much as possible for intuition, assume we have a sample with no variation in p, z or d,

and no measurement error, so $e = 1$. The private assignable demand functions are then

$$y_j = h_j(\eta_j(e)\,x) \quad \text{for } j = 1, ..., J. \tag{5}$$

Here e is a realization of the random, unobserved resource share vector V, and y_j is the corresponding realization of a random quantity demand Y_j. For simplicity just write $\eta_j(V)$ as η_j remembering now that η_j is a random variable, not a constant. From data we can identify $F_Y(y_1, ..., y_J \mid x)$, that is, the joint distribution of Y_1,....Y_J, conditional on x. The goal is to identify the distribution of the random vector η_1,...,η_{J-1}. Again since resource shares sum to one, if that distribution is known, then we also know the distribution of the last resource share η_J.

What we show in our theorems is that knowledge of $F_Y(y_1, ..., y_J \mid x)$ is sufficient to back out the joint distribution of η_1,...,η_{J-1} up to an unknown location for each η_j. For example, just from F_Y we could not determine the mean of each η_j, but we can determine its distribution around that unknown mean. The key element to showing this is that the unobserved V act on observed demands only through the resource shares η_j. If the functions h_j were known, we could invert them to get $h_j^{-1}(Y_j) = \eta_j$, and the distribution of η_j would be identified from the distribution of $h_j^{-1}(Y_j)$.

The key to our identification theorem is showing that the h_j^{-1} function can be identified up to an unknown location, thereby letting us identify the distribution of the resource shares up to an unknown location for each share. We can then combine this identification with the results from the previous subsection (or with DLP or BCL) to get the unknown locations for each η_j, and thereby identify their entire joint distribution.

To see why it is possible to obtain the h_j^{-1} function up to an arbitrary location, consider the probability that $Y_j \leq y$, conditional on x, for any given values of y and x. That probability is something we can estimate from data, and by the model that probability equals the probability that $h_j(\eta_j x) \leq y$. For any change in y, we can calculate the corresponding change in x that leaves this probability unchanged and hence leaves the value of $h_j(\eta_j x)$ unchanged. This allows us to trace out how the function $h_j(\eta_j x)$ changes with x, identifying h_j^{-1} function up to an arbitrary location. As in the previous subsection, in our real theorems we need to include z, d (if any) and p (if the data contain price variation) in the derivations.

Theorem 1 later formalizes this identification around unknown location, and Theorem 2 then adds the earlier described identification of location based on distribution factors. Using a very different proof technique, Chiappori and Kim (2017) also obtain an identification result for random resource shares around unknown locations, but their result requires substantial price variation, deals only with two-member households and has no measurement error e.

So, e.g., their results could not be applied either to Engel curve data, or to a model with children.

3.5 Some Details

Here we give a few more details regarding distribution identification. Formal theorems and derivations are in the next subsection and the appendix.

For exposition, focus on a single j, and continue to drop p, d, z and e. Let $F_{Y_j}(y_j \mid x)$ be the distribution function of Y_j given x that we can identify and estimate from data, and let $F_{\eta j}(\eta_j)$ denote the unknown distribution of the random resource share η_j. Assume the function h_j is invertible. For any given value y_j and x we have that

$$\begin{aligned} F_{Y_j}(y_j \mid x) &= \Pr(Y_j \leq y_j \mid x) = \Pr(h_j(\eta_j x) \leq y_j \mid x) \\ &= \Pr\left(\eta_j x \leq h_j^{-1}(y_j) \mid x\right) = \Pr\left(\eta_j \leq \frac{h_j^{-1}(y_j)}{x}\right) = F_{\eta j}\left(\frac{h_j^{-1}(y_j)}{x}\right) \end{aligned}$$

where the last equality uses the assumption that the distribution of V and hence η_j is independent of x. Let $\gamma = h_j^{-1}(y_j)/x$. We want to use this expression to learn about the h_j function, but we don't know $F_{\eta j}$. However, we can use the above to get an expression only involving h_j. Consider the ratio of derivatives

$$\begin{aligned} \frac{-\partial F_{Y_j}(y_j \mid x)/\partial y_j}{\partial F_{Y_j}(y_j \mid x)/\partial \ln x} &= \frac{-\partial F_{\eta j}\left(h_j^{-1}(y_j)/x\right)/\partial y_j}{x\partial F_{\eta j}\left(h_j^{-1}(y_j)/x\right)/\partial x} = \frac{-F'_{\eta j}\left(h_j^{-1}(y_j)/x\right)\frac{h_j^{-1\prime}(y_j)}{x}}{xF'_{\eta j}\left(h_j^{-1}(y_j)/x\right)\frac{h_j^{-1}(y_j)}{-x^2}} \\ &= \frac{h_j^{-1\prime}(y_j)}{h_j^{-1}(y_j)} = \frac{\partial \ln h_j^{-1}(y_j)}{\partial y_j} \end{aligned}$$

where the apostrophe denotes differentiation. Note the second equality applies the chain rule to both the numerator and the denominator in the ratio of derivatives. The function $F_{\eta j}$ drops out, leaving the log derivative of h_j on the right-hand side. The left-hand side of this equation is something identified (that we can estimate from data), so this log derivative of h_j^{-1} is identified. Integrating this expression with respect to y_j then identifies the function h_j itself, up to an unknown constant of integration. This unknown constant of integration is the unknown location or mean term discussed earlier.

In our data we have J quantity demands $y_1, ..., y_J$ to work with, but we only needed to identify the distribution of $J-1$ resource shares $\eta_1, ..., \eta_{J-1}$, because the last one η_J is obtained from the others by the constraint that resource shares sum to one. This gives us one more demand equation to work with than we need for Theorem 1, which we can use to

identify one more dimension of random variation. That is where the measurement error e comes in. We assume the distribution of e is independent of the true resource shares, which is reasonable since e is measurement error and the variation in resource shares corresponds to true behavioural heterogeneity across households. With measurement error e, Theorem 1 uses a technique like that given above to identify the joint distribution of $\eta_1 e$,....,$\eta_J e$ up to unknown locations (means), and Theorem 2 identifies those locations. We can then use $\sum_{j=1}^{J} \eta_j e = e$ to identify the distribution of e, and separate that out from the joint distribution of the resource shares η_1,....,η_{J-1}.

3.6 Formal Identification

In this section we provide formal assumptions and our associated identification theorems. Proofs are in the appendix.

ASSUMPTION A1: For every individual $j \in \{1, ..., J\}$ in the household there is a private, assignable good, denoted as good j. The household's demand function for good j is given by $Y_j = h_j(\eta_j e x, p, z)$. The unknown functions $h_1, ..., h_J$ are differentiable and strictly increasing in their first element. Resource shares η_1,....,η_J are random variables having some unknown joint distribution across households, with $\sum_{j=1}^{J} \eta_j = 1$. The measurement error e is also a random variable that has an unknown distribution across households.

As discussed earlier, the assumptions that $Y_j = h_j(\eta_j e x, p, z)$ with $\sum_{j=1}^{J} \eta_j = 1$ follow immediately from the general framework of the Pareto efficient collective household model, with the goods indexed by j being private and assignable (other goods may be shared and non-assignable). The assumption that each h_j is increasing in its first element just means that good j is a normal good, i.e., a good for which demand goes up when total expenditures goes up.

Distribution factors d are defined to be characteristics that affect η_j but not h_j. Previous collective household models assumed that each η_j is a deterministic function of d and other observed variables. In contrast, we assume that η_j varies randomly across households, or equivalently, that there exist unobserved distribution factors. This variation in each η_j induces variation in observed private assignable goods demands Y_j.

Let $F_Y(Y_1, ...Y_j \mid p, x, d, z)$ denote the joint distribution of expenditures on private assignable goods $Y_1, ...Y_j$, conditioning on p, x, d, and z.

ASSUMPTION A2: $F_Y(Y_1, ...Y_J \mid p, x, d, z)$ is identified from data for all p, x, d and z in some sets Φ_p, Φ_x, Φ_d and Φ_z, respectively. The set Φ_x is an interval, and the set Φ_p is not empty.

A sufficient but stronger than necessary condition for Assumption A2 to hold is that we have *iid* observations of standard household survey data $(Y_1, ...Y_J, p, x, d, z)$ where the support of p, x, d, z includes $\Phi_p \times \Phi_x \times \Phi_d \times \Phi_z$. But Assumption A2 could alternatively hold under many other standard sampling schemes, such as repeated cross-sectional data where prices only vary by time, or situations where there is weak dependence in the observations, such as mixing. The set Φ_z could be empty, corresponding to not observing any demographic or other characteristics z that might affect tastes. The set Φ_p could just contain a single element, in which case we will have Engel curve data with no price variation, as in a single cross section of households. Having Φ_x contain an interval implies that observed households have some continuous variation in total expenditures x, though the range of observed total expenditure values x, given by the interval Φ_x, could be very small.

ASSUMPTION A3: Assume that η_1,...,η_J, conditional on any $p \in \Phi_p$, $d \in \Phi_d$, $z \in \Phi_z$, $x \in \Phi_x$, and any e, has a continuous distribution that is independent of x and e. Let $F_\eta(\eta_1, ..., \eta_J \mid p, d, z)$ denote the unknown joint distribution of η_1,...,η_J conditional on p, d, z.

Lewbel and Pendakur (2008) and Bargain and Donni (2009) assume $e = 1$ and η_1,...,η_J independent of x as in Assumption A3 to obtain identification in the case of deterministic rather than random resource shares. BCL and Lise and Seitz (2011) imposed η_1,...,η_J independent of x on their empirical models (again having deterministic shares). DLP and Menon, Pendakur, and Perali (2012) provide both theoretical and empirical evidence supporting this independence assumption, and Cherchye, De Rock, Lewbel and Vermeulen (2015) find that identified bounds on resource shares also support this assumption. Note that this independence only needs to hold after conditioning on observables z that can include demographic characteristics and observable distribution factors. One way to interpret Assumption A3 is to assume that unobservable distribution factors exist, including at least one that is continuously distributed, and that the distribution of unobserved distribution factors across households does not depend on x, after conditioning on p and z.

ASSUMPTION A4: Assume that $e > 0$ is independent of η_1,...,η_J,p, d ,x, and z, that $E(e) = 1$, and that $\eta_1, ..., \eta_{J-1}$,e has a nonvanishing characteristic function.

The assumption that $E(e) = 1$ means that $E(x^*) = E(x)$, corresponding to the standard measurement error assumption that the observed variable is on average correctly measured, so total expenditures are not systematically over or undercounted.

THEOREM 1: Let Assumptions A1, A2, A3 and A4 hold. Then, for some unknown

functions $c_1(p,z),...,c_J(p,z)$ the joint distribution of $\eta_1 e c_1(p,z),...,\eta_J e c_J(p,z)$ conditional on p,d,z is identified for all $p \in \Phi_p$, $d \in \Phi_d$ and $z \in \Phi_z$.

A well-known nonidentification result in the collective household literature is that without restrictions on preferences, the levels of (deterministic) resource shares cannot be identified (see Chiappori and Ekelund 2009 for a current general version of this result). Instead, only changes in resource shares with respect to observed distribution factors can be identified. This is equivalent to saying that, if e and each η_J were deterministic functions of observables, then only $\eta_1 e C_1(p,z),...,\eta_J e C_J(p,z)$ could be identified for some unknown functions $C_j(p,z)$. Theorem 1 provides a substantial generalization of this result to random resource shares. For example, consider the special case where e is constant. Theorem 1 then says that when η_j are random, the entire joint distribution of resource shares (and hence the effect of all unobserved distribution factors) is identified up to the same unknown deterministic functions $C_j(p,z) = ec_j(p,z)$. The distribution identification result in Chiappori and Kim (2017) is also a special case of Theorem 1, and in particular follows from Theorem 1 when e is constant and $J = 2$. Note that Chiappori and Kim also require considerable price variation, while our results allow Φ_p to have few or even just one element.

ASSUMPTION A5: Assume Φ_d contains at least J elements, which without loss of generality will be denoted $d_1,...,d_J$. Assume $E(\eta_j \mid p,d,z) \neq 0$. Let $T(p,z)$ be the J by J matrix defined by having $E(\eta_j \mid p,d_k,z)/E(\eta_j \mid p,d_1,z)$ in the row k and column j position. Assume there exist sets Φ_p^* and Φ_z^*, which are subsets of Φ_p and Φ_z, such that for all $p \in \Phi_p^*$ and $z \in \Phi_z^*$, $T(p,z)$ is nonsingular.

The key feature of Assumption A5 is the requirement that our set of distribution factors must take on at least J values, recalling that J is the number of household members. The nonsingularity of $T(p,z)$ required by Assumption A4 will generally hold, failing only when there is some equality coincidence among the expected resource share functions $E(\eta_j \mid p,d_1,z)$. For example, in households with two members, it is straightforward to check that nonsingularity will hold as long the distribution factor affects the mean of η_1 in any way, that is, as long as $E(\eta_j \mid p,d_1,z) \neq E(\eta_j \mid p,d_2,z)$. Also, regarding $E(\eta_j \mid p,d,z) \neq 0$, we actually only require that this hold for one tuple of given known, observed values of p, d and z.

THEOREM 2: Let Assumptions A1, A2, A3, A4 and A5 hold. Then $F_\eta(\eta_1,...,\eta_J \mid p,d,z)$ is identified for all $d \in \Phi_d$, $p \in \Phi_p^*$ and $z \in \Phi_z^*$. The distribution of e is also identified.

Theorem 2 works by exploiting the fact that resource shares must sum to one within a household. For example, if Φ_d has J elements, this places J equality constraints on the set

of functions $E(\eta_j \mid p, d, z)$, one for each of the J values that the distribution factors can take on. By Theorem 2, the $E(\eta_j \mid p, d, z)$ are identified up to J unknown functions, and the J equality constraints allow us to recover these J unknown functions, $c_1(p, z), ..., c_J(p, z)$. Given these $c_j(p, z)$ functions, by Theorem 2 the entire joint distribution of the resource shares is identified.

Note that Theorems 1 and 2 do not require any price variation, and so can be applied to Engel-curve-type data where all observations are drawn from a single price regime.

An immediate extension of Theorem 2 is that we could have identified the levels of deterministic resource shares, e.g., in traditional nonstochastic collective household models, given resource shares independent of total expenditures and observing some distribution factors. Theorem 2 thereby overcomes the classic resource shares nonidentification problem. So, e.g., the SAP and SAT preference restrictions employed by DLP could be replaced with Assumption A4.

4 Semiparametric Estimation

Although we have established nonparametric identification of the model, completely nonparametric estimation is likely to be infeasible with realistically sized data sets, given the high dimensionality of the model. We therefore focus on estimation assuming a parametric specification for the private assignable demand equations and for the deterministic component of resource shares, but a nonparametric specification for the random variation in resource shares. From the long literature on empirical demand estimation, a great deal is known about functional forms for demand functions that are empirically successful. However, almost nothing is known about the distribution of random variation in resource shares. We therefore employ commonly used functional forms for utility, while leaving the distribution of resource shares (and measurement error e) unspecified and nonparametric.

In our application we take men's, women's and children's clothes to be our private assignable goods y_j. Let θ_j be a vector of clothing demand parameters. Other variables are defined as before. Based on our theorems, assume $(\eta_1, \eta_2, \eta_3, e) \perp x \mid z, p, d$ and let $E(e \mid \eta_1, \eta_2, \eta_3, z, p, d) = 1$. The general form of a demand function is

$$y_j = h_j(x\eta_j e, p, z \mid \theta_j)$$

where the distribution of η_j can depend on z, p, d. In the appendix we provide a general technique based on inverse demands for estimating θ_j and the values of $\eta_1, \eta_2, \eta_3, e$ for each individual in the sample. Below we give a simpler estimator based on our specifically chosen

functional form for h_j.

Our empirical application uses cross-sectional data from a single time period, so we assume that p is fixed; i.e., we work with Engel curve data. We can then drop p from the above demand equation, and let each y_j just equal observed expenditures on member j's clothing. The estimator described below will allow us to directly estimate the mean value of $(\eta_1, \eta_2, \eta_3, e)$, conditional on covariates, and then later back out estimated values of η_1, η_2, η_3, and e for each individual in each household in our sample.

We assume that the Marshallian demand function h_j for the private assignable good has the commonly used piglog (price independent generalized logarithmic) functional form. The name piglog comes from Muellbauer (1976). This is one of the most widely used specifications for Engel curves, e.g., the Jorgenson, Lau, and Stoker (1982) Translog demand system and the Deaton and Muellbauer (1980) Almost Ideal Demand System have Engel curves of the piglog form, and piglog Engel curves were also used in empirical collective household models estimates by DLP and by Calvi (2017). Note that only the assignable goods are assumed to be piglog, so more general specifications could apply to other goods, such as the quadratic almost ideal demand system (QUAIDS) model of Banks, Blundell, and Lewbel (1997).

The piglog Engel curve model is[3]

$$h_j(x, z \mid \theta_j) = c_j(z)\, x + b_j(z)\, x \ln(x), \tag{6}$$

where $c_j(z)$ and $b_j(z)$ are implicitly functions of parameters θ_j. The corresponding household demand function for each private assignable good j is

$$y_j = c_j(z)\, x\eta_j e + b_j(z)\, x\eta_j e \ln(x\eta_j e).$$

Let $w_j = y_j/x$ be the budget share of person j's assignable good as a fraction of total household expenditures x. We then have that the household's budget-share function for the assignable good j is

$$w_j = [c_j(z) + b_j(z) \ln(x\eta_j e)]\, \eta_j e.$$

[3]Piglog demands are given by the class of indirect utility functions $V(\ln P, \ln x, z) = (\ln x - A(P, z)) \exp(B(P, z))$ for some functions $A(P, z)$ and $B(P, z)$, where $\exp(A(P, z))$ is homogeneous of degree one in P, and $B(P, z)$ is homogeneous of degree zero in P. Applying Roy's identity to this indirect utility gives the piglog demand functions, in which the budget share of good j equals $A_j(P, z) - A(P, z) B_j(P, z) - B_j(P, z) \ln x$, where $A_j(P, z)$ is the partial derivative of $A(P, z)$ with respect the price of good j, and similarly for $B_j(P, z)$. Holding prices fixed at some constant level P_0, define $b_j(z) = -B_j(P_0, z)$ and $c_j(z) = A_j(P_0, z) - A(P_0, z) B_j(P_0, z)$ to obtain equation (6).

Define μ_j to be the conditional mean function for η_j, given by

$$\mu_j(d,z) = E(\eta_j e \mid x,d,z) = E(\eta_j \mid d,z).$$

Define U_j to be a structural error term, which depends on the deviation of η_j from μ_j as

$$U_j = \frac{\eta_j e}{\mu_j(d,z)}.$$

If resource shares η_j are deterministic functions of observables, then $\mu_j(d,z) = \eta_j$, so in that case μ_j are the resource shares. Otherwise, if resource shares η_j are random (functions of unobserved distribution factors e) then U_j is the random variation in $\eta_j e$ across individuals. Substituting these expressions into the household budget-share functions gives

$$w_j = (c_j(z) + b_j(z)(\ln(x\mu_j(d,z)) + \ln U_j))\,\mu_j(d,z)\,U_j.$$

Gathering U_j terms this becomes

$$w_j = [c_j(z) + b_j(z)\ln(x\mu_j(d,z))]\,\mu_j(d,z)\,U_j + b_j(z)\,\mu_j(d,z)\,U_j \ln U_j. \tag{7}$$

Since $E(e \mid \eta_1,\eta_2,\eta_3,z,p,d) = 1$ and $E(\eta_j \mid d,z) = \mu_j(d,z)$, we have that $E(U_j \mid x,d,z) = 1$. Let $s_j = E(U_j \ln U_j \mid x,d,z)$, which for simplicity is assumed to be constant (letting s_j depend on covariates is feasible, but complicates the numerical evaluations below). This is roughly analogous to a simplifying homoscedasticity-type assumption. Define an additive error term ε_j by $\varepsilon_j = w_j - E(w_j \mid x,d,z)$, and therefore $E(\varepsilon_j \mid x,d,z) = 0$. Evaluating $E(w_j \mid x,d,z)$ using the above expression for w_j gives

$$w_j = [C_j(z) + b_j(z)\ln(x\mu_j(d,z))]\,\mu_j(d,z) + \varepsilon_j, \tag{8}$$

where $C_j(z) = c_j(z) + b_j(z)\,s_j$. We can think of the U_j's as structural errors, while the ε_j's are ordinary regression errors.

We will later assume for estimation that $C_j(z)$, $b_j(z)$ and $\mu_j(d,z)$ are all linear in their arguments, so

$$\mu_j(d,z) = \alpha_{j0} + \alpha_{jz}'z + \alpha_{jd}'d, \quad b_j(z) = \beta_{j0} + \beta_{jz}'z \quad \text{and} \quad C_j(z) = \gamma_{j0} + \gamma_{jz}'z$$

for constant scalars α_{j0}, β_{j0} and γ_{j0}, and constant coefficient vectors, α_{jz}, α_{jd}, β_{jz} and γ_{jz}. Substituting these expressions into equation (8) gives the model that we estimate, which for

each member j of each household h is

$$w_{jh} = \left[\gamma_{j0} + \gamma'_{jz} z_h + \left(\beta_{j0} + \beta'_{jz} z_h\right) \left[\ln x_h + \ln\left(\alpha_{j0} + \alpha'_{jz} z_h + \alpha'_{jd} d_h\right)\right]\right]\left(\alpha_{j0} + \alpha'_{jz} z_h + \alpha'_{jd} d_h\right) + \varepsilon_{jh}. \tag{9}$$

Equation (9) for household members $j = 1, \ldots, J$ now forms a system of J equations that we can estimate using a least squares estimator such as nonlinear seemingly unrelated regression. We will also want to employ clustered (across equations), heteroskedastic robust standard errors, to account for the fact that, by construction, the errors ε_j are heteroskedastic and are correlated across the members of each household.

In our empirical work, due to data limitations we assume a single utility function for all the children in the household. This utility function includes a children's assignable good, which is children's clothing and shoes (DLP make the same assumption for the same reason). If we had observed separate clothing or shoes expenditures for each child, we could have allowed each to have their own utility function. Let $j = children$ index the children's aggregate resource share and their aggregate clothing budget share. For simplicity in reporting the estimated results, assume the children split their resource shares equally. As a result, the children's aggregate clothing budget share equation is given by equation (9), except that we replace $\ln x_h$ with $\ln\left(x_h/s\right)$ where s is the number of children in the household.

An interesting feature of this model and our associated estimator is that the estimating equations (9) for each j are the same regardless of whether resource shares are random or not. If resource shares are not random then $s_j = 0$ making $c_j(z) = C_j(z)$ and $\eta_{jh} = \mu_j(d_h, z_h)$, where η_{jh} is the resource share of member j in household h. But if resource shares are random, then (given the parameter estimates) we can use the fact that the errors ε_{jh} are known functions of the random components U_{jh} to recover random resource shares for each household, as follows.

By construction, we have that for each household member j of each household h,

$$\varepsilon_{jh} = w_{jh} - E\left(w_{jh} \mid x_h, d_h, z_h\right) = \left(U_{jh} - 1\right)\psi_{0jh} + \left(\ln U_{jh} - s_j\right) U_{jh}\psi_{1jh}, \tag{10}$$

where ψ_{0jh} and ψ_{1jh} are known functions of our estimated parameters. Specifically, from equations (8) and (7),

$$\begin{aligned} \psi_{0jh} &= \left[C'_j(z_h) + b_j(z_h)\ln\left(x\mu_j(d_h, z_h)\right)\right]\mu_j(d_h, z_h) \\ \psi_{1jh} &= b_j(z_h)\,\mu_j(d_h, z_h). \end{aligned}$$

So for each member j of each household h, given the residual ε_{jh} and our estimates of ψ_{0jh} and ψ_{1jh}, if we knew s_j we could numerically solve (10) for U_{jh}. By iterating over the sample

averages of $U_{jh} \ln U_{jh}$ and plugging these in for s_j in (10), we arrive at sample values of U_{jh} consistent with the observed error terms z_{jh} and the model. Note that equation (10) may have up to two solutions U_{jh}. We use only strictly positive solutions for U_{jh}, and where there are two strictly positive solutions, we use the one that gives us sample means closest to one (since the true mean of U_{jh} is one).[4]

Once we have obtained each person's U_{jh}, we then calculate

$$c_h \quad \sum_{j=1}^{J} U_{jh} \mu_j (d_h, z_h) \quad \text{and} \quad \eta_{jh} \quad \frac{U_{jh} \mu_j (d_h, z_h)}{c_h}$$

where c_h is the multiplicative measurement error in x_h for household h and η_{jh} is the resource share of person j in household h. These expressions follow from the constructions $U_{jh} = \eta_{jh} c_h / \mu_j (d_h, z_h)$ and $\sum_{j=1}^{J} \eta_{jh} = 1$.

5 Empirical Results

5.1 Malawian Expenditure Data

We use almost the same Malawian household expenditure and demographic data as DLP, supplemented with some additional distribution factor variables.[5] Malawi is a very poor former British protectorate in southern Africa. With a population density of approximately 190 persons per square kilometre, it is one of the most densely populated countries in Africa, even though roughly 83 per cent of the population lives in rural areas. Unlike many poor African countries, Malawi's recent political history is remarkable for the absence of military coups and for occasional multi-party elections, most recently in 2009. Despite its relative political stability, Malawi has numerous socio-economic tensions, including extreme poverty, a high incidence of HIV/AIDs, high infant mortality and one of the lowest life expectancies in the world (51 years).

While household-level poverty rates are quite high, we find that accounting for intra-

[4] We could impose the restriction that the mean of U_{jh} is exactly one by implementing a GMM estimator that includes the moment conditions based on $E[(U_{jh} \ln U_{jh} - s_j) \mid x, d, z] = 0$ and $E[(U_{jh} - 1) \mid x, d, z] = 0$, where U_{jh} is defined implicitly by equation (10). We instead just take s_j to be the sample mean of $U_{jh} \ln U_{jh}$ and check that the sample mean of U_{jh} is close to one, for each value of j.

[5] The data used in this paper are from the second Malawi Integrated Household Survey (IHS2), conducted in 2004 – 2005, and jointly designed by the National Statistical Office of the Government of Malawi (NSOM) and the International Food Policy Research Institute and the World Bank (http://www.nso.malawi.net (accessed June 2007)). The survey includes roughly 11,000 households. The data are confidential and were released under agreement with the NSOM. Below we discuss the small differences between our data selection and DLP.

household allocations results in even higher rates of poverty. In particular, our results show many people (particularly children) are individually below a poverty line, despite living in households where everyone in the household would be above the line if intra-household resources were distributed more equitably.

These are relatively high-quality data. Enumerators were sent to individual households to collect the data for IHS2. Enumerators were monitored by field supervisors in order to ensure that the random samples were followed and also to ensure data quality. Cash bonuses, equivalent to roughly 30 per cent of average household income in Malawi, were used as an incentive system in the IHS2 for all levels of workers. Roughly 5 per cent of the original random sample was resampled because dwellings were unoccupied. Only 0.4 per cent of initial respondents refused to answer the survey in the IHS2, so endogenous selection of reporters is not likely to be a problem in these data.

In the survey, households are asked questions from a number of modules relating to health, education, employment, fertility and consumption. Households are asked to recall their food consumption (one week recall) and their non-food expenditures broken into four recall categories (one week, one month, three months and one year). In our work, we convert consumption to the annual level (e.g., multiplying one month recall consumption by 12). Consumption amounts also include the value of home-produced goods and services imputed at the value of those services consumed in the market.

The consumption data include (in the three-month recall questionnaire) household expenditures on clothing and shoes for the household head, spouse(s), and children aged 14 and under. These are the assignable goods whose demand functions we model. We use a single private assignable good for each person equal to the sum of clothing and footwear expenditures for that person. As noted above, we treat "children" as an aggregate person, and interpret that resource share as being divided equally among the children in the household.

Our new method for identifying the level of resource shares depends on observing distribution factors, defined as variables that affect resource shares but do not affect individual preferences. We construct distribution factors based on the following variables: (i) indicators that the husband and wife live in their home village; (ii) the natural logarithm of one plus the distances to the nearest doctor, the nearest bank, the nearest post-office, the nearest weekly market and the nearest government high school in kilometres; (iii) self-reported quality of local elementary and secondary school, on a scale from -0.2 to 0.6, where 0 is the sample mode; (iv) an indicator that the household has experienced a climate shock in the preceding 4 years that affected agricultural productivity; (v) the natural logarithm of the local mill price in Malawian kwacha, reflecting the value of home production. We note that only (i), (iii) and (iv) are individually varying variables; all distances and the local mill price are

community-level variables. To mitigate multicollinearity issues, we extract the first three principal components of the above list of variables and use these principal components as the distribution factors d in our structural models. Note that our primary interest is in using variation in d for resource share identification, rather than uncovering the effects of any of these specific variables on resource shares.[6]

We include some demographic variables, denoted z, which may affect both preferences and resource shares of each household member. We include 17 demographic variables in our models: region of residence (non-urban North and non-urban Central with non-urban South as the left-out category); the average age of children less 5; the minimum age of children less 5; the proportion of children who are girls; the age of the man less 28 and the age of the woman less 22 (the average ages of men and women in the sample); the education levels of the household head and spouse (ranging from -2 to 4, where 0 is the modal education level); the log of the distance of the village to a road and to a daily market; a dummy indicating that the three-month recall period for consumption occurred over the dry season; and dummy variables indicating that the household is Christian or Muslim (with animist/other as the left-out category). Additionally, we include the number of children, s, among our demographic variables, in the form of separate dummy variable indicators for two-, three- and four-child households. These are the same demographic variables as used in DLP. Thus, z has 17 elements: 14 demographic elements and 3 elements equal to values of s.

We use almost the same sample restrictions as DLP. As in that paper, our sample consists of married couples with one to four children all under 15 years of age that satisfy the following sample restrictions: (i) polygamous marriages are excluded; (ii) observations with any missing data on the age or education of members are excluded; (iii) households with children aged 15 or over are excluded; (iv) households with any member over 65 are excluded; and (v) urban households are excluded. In the current paper, though, we additionally exclude households with servants (dropped 78 households) and include households that have missing data not used in the estimation, such as wealth variables used by DLP (added 89 households). Further, we dropped any household with missing values for any of the above distribution factors (dropped 206 households).

Our final sample contains 2,599 households. Table 1 gives some summary statistics.

[6] If desired, factor loadings on these variables could be used to uncover the effect of any particular variable (through the distribution factor principal components) on resource shares and within household inequality.

Table 1: Summary Statistics

		mean	std dev	min	max
clothing	man	0.013	0.026	0.000	0.529
shares	woman	0.019	0.026	0.000	0.479
	children	0.015	0.020	0.000	0.261
(left-out)	$s = 1$	0.306	0.461	0.000	1.000
	$s = 2$	0.302	0.459	0.000	1.000
	$s = 3$	0.231	0.422	0.000	1.000
	$s = 4$	0.161	0.367	0.000	1.000
principal	first	0.000	1.373	-4.147	3.767
components	second	0.000	1.096	-3.596	3.071
of dist. factors	third	0.000	1.051	-6.473	2.727

5.2 Estimation

We estimate equation (9) by nonlinear seemingly unrelated least squares. To account for variation in the number of children across households, in equation (9) we replace $\ln x_h$ in the children's assignable-good equation with $\ln(x_h/s)$. The three distribution factors d are included in all models.

To reduce the complexity and potential multicollinearity of the nonlinear estimation, we impose the restriction that, while the three dummy variables for two-, three- and four-child households ($s = 2, 3, 4$) enter both the slope terms $\beta'_{jz} z_h$ and the level terms $\gamma'_{jz} z_h$, the other 14 demographic variables enter only through the level terms $\gamma'_{jz} z_h$. Thus, we hereafter refer to β_{js} as budget response parameters for the three types of people j (man, woman, children) and three household sizes ($s = 2, 3, 4$), while β_{j0} refers to the baseline budget response (equal to the budget response for the left-out household size where $s = 1$).

We provide estimates where identification hinges on the existence of observed distribution factors as in Theorem 1. We also provide estimates imposing the similar across people (SAP) restrictions of DLP. Note in particular that SAP imposes the constraint that β_{js} does not vary across household members, so $\beta_{js} = \beta_s$ for $s = 0, 2, 3, 4$.

For the model identified using observed distribution factors, the sum-of-squared residuals function has many local minima in our empirical context, regardless of the set of distribution factors chosen. To circumvent difficulties with local minima, we perform a grid search over 1,250 random starting values for η and β and select the estimates corresponding to the minimum of the observed sum-of-squared residuals. All estimates employ this grid search strategy.[7]

Unlike identification based on distribution factors, identification using the SAP restriction

[7] We used Stata 14 on a 600 node Linux cluster to compute the estimates, using five different seeds and drawing 250 starting values per seed.

does not suffer from the problem of multiple local minima in the objective function. DLP note similar results when just comparing SAT and SAP, noting that the SAP restriction facilitates identification of resource share parameters in particular.

The next subsection reports our estimates of the levels of resource shares. This is then followed by estimates of the random distribution of resource shares around these levels, and then the resulting estimates of separate povery rates for men, women and children.

5.3 Estimated Coefficients and Levels of Resource Shares

Table 2 gives estimates of selected elements of β_{js} and of the resource share parameters α for our model, and for the more restrictive SAP model of DLP (Table A1 in the Appendix provides the full set of estimates of all parameters). At the bottom of the SAP block, we provide a test of these restrictions $\beta_{js} = \beta_s$, which is a test of SAP subject to the validity of the distribution factors.

In the upper panel of the leftmost block, we show selected estimates of resource share parameters where identification comes from the three observed distribution factors, denoted DF. Consider first the estimated values for the number of children. The upper left coefficient is $\alpha_{man,0}$, which gives the estimated level of the resource share for a man in a household with one child and all other demographic variables and distribution factors equal to zero. This indicates that the man in such a household has a shadow budget equal to roughly 34.7 per cent of the overall household budget. The estimate of 0.392 for $\alpha_{woman,0}$ indicates that the woman in such a household has a shadow budget equal to about 39.2 per cent of the overall household budget. This leaves 26.1 per cent of the household budget for the child. Values for the children's resource share parameters are thus calculated rather than estimated, and their standard errors are obtained via the delta method.

Despite differences in our sample regressors, and particularly in the identification strategy relative to DLP, our estimated level of resource shares of men, women and children are similar to that earlier work. In particular, as the number of children in the household goes up, the man's resource share doesn't change, the woman's resource share goes down, and the childrens' (aggregate) resource share goes up. The sample value of the χ^2_3 test statistic for the Wald test of the hypothesis that the man's resource share does not respond to the number of children is 0.87, which does not reject at conventional levels. In contrast, the sample values of the test statistics for the analogous restriction on the women's and children's shares are 15.16 and 9.10, respectively, which do reject at conventional levels.

Turning to the distribution factors, we find that they are in general somewhat weak in driving resource shares. In particular, no distribution factor is statistically significant at the

Table 2: Selected Parameter Estimates

		DF		SAP	
		est	*std err*	est	*std err*
α_{man}	constant	0.347	*0.058*	0.316	*0.054*
	$s = 2$	0.009	*0.031*	0.016	*0.026*
	$s = 3$	0.028	*0.032*	0.031	*0.028*
	$s = 4$	0.015	*0.044*	0.014	*0.034*
	d_1	0.007	*0.005*	0.007	*0.004*
	d_2	-0.007	*0.004*	-0.007	*0.004*
	d_3	0.005	*0.005*	0.004	*0.005*
α_{woman}	constant	0.392	*0.098*	0.396	*0.049*
	$s = 2$	-0.065	*0.021*	-0.074	*0.022*
	$s = 3$	-0.072	*0.023*	-0.080	*0.023*
	$s = 4$	-0.095	*0.027*	-0.102	*0.027*
	d_1	-0.007	*0.004*	-0.007	*0.004*
	d_2	0.002	*0.003*	0.002	*0.003*
	d_3	-0.002	*0.004*	-0.001	*0.004*
$\alpha_{children}$	constant	0.261	*0.092*	0.288	*0.042*
	$s = 2$	0.056	*0.020*	0.058	*0.016*
	$s = 3$	0.044	*0.018*	0.048	*0.015*
	$s = 4$	0.080	*0.029*	0.088	*0.021*
	d_1	0.000	*0.003*	0.000	*0.003*
	d_2	0.005	*0.004*	0.005	*0.003*
	d_3	-0.002	*0.004*	-0.003	*0.004*
β_{man}	constant	0.016	*0.003*	0.018	*0.003*
	$s = 2$	0.005	*0.003*	0.005	*0.004*
	$s = 3$	0.006	*0.004*	0.007	*0.004*
	$s = 4$	0.002	*0.004*	0.000	*0.004*
β_{woman}	constant	0.023	*0.009*		
	$s = 2$	-0.001	*0.008*		
	$s = 3$	0.006	*0.008*		
	$s = 4$	-0.011	*0.010*		
$\beta_{children}$	constant	0.013	*0.007*		
	$s = 2$	0.021	*0.011*		
	$s = 3$	0.014	*0.010*		
	$s = 4$	0.009	*0.011*		
test, p-val				7.97	*0.436*

5 per cent level, and only three are statistically significant at the 10 per cent level (the first and second principal component for men, and the first principal component for women). The test that all three distribution factors can be excluded from the resource shares is 6.64 and 4.08 for the man's and woman's resource share, respectively, indicating marginal statistical significance for men only. The main reason for the relatively low significance of the distribution factors appears to be a multicollinearity problem, in that these factors are highly correlated with our observed demographic preference shifters.

Next consider the slope parameters β_{js}. We find that the estimated values of the parameters β_{j0} are all positive and statistically significant (nonzero slope coefficients are required for our resource share identification). Positive β_{j0} indicates that clothing is a luxury good in the Malawian context. We find that the values of $\beta_{man,0}$, $\beta_{woman,0}$ and $\beta_{children,0}$ vary somewhat across people; e.g., the estimated value for women is 0.023 and for men is 0.016. However, these differences across people are not statistically significant. In particular the sample value of the χ^2_8 test statistic for the hypothesis that the man, woman and children have the same budget responses at each household size is 7.97, which has a p-value of 0.436. Thus, we cannot reject the restriction that preferences are similar across people (SAP).

Some of the resource share estimates described above have relatively large standard errors, due to the weakness of the distribution factors noted above. However, since the SAP restrictions are not rejected, we can consider imposing them to increase estimation precision, noting that SAP estimation did not suffer from the numerical problem of multiple minima.

The right-hand panel of Table 2 shows the estimated values of resource shares, distribution factor coefficients, and β_s coefficients given SAP. The resource shares are quite similar to the estimates that didn't impose SAP, but imposing SAP substantially reduces standard errors. Even with SAP, the distribution factor coefficients remain mostly insignificant.

Let $\mu_{jh} = \mu_j(d_h, z_h)$. Table 3 summarizes the variation in μ_{jh}, that is, the variation in estimated resource shares due only to variation in observed demographic variables z and distribution factors d.

For men, women and children, we give the mean and standard deviations (across z and d values) of μ_{jh} in the row labelled "all." The mean and standard deviation of μ_{jh} just due to variation in distribution factors is given in the row denoted "d." For men, the average resource share is 0.468 (the average contribution from observed distribution factors is zero because by construction the distribution factors themselves have an average of zero). The standard deviation of resource shares for men is 0.155, and that of the contribution from distribution factors is 0.013, about a tenth as large. We find similar relative contributions of resource shares for women and children.

The lower block in each panel gives the average and standard deviation of resource shares

Table 3: Summary Statistics for Estimated Resource Shares

		DF		SAP	
2599 observations		mean	std dev	est	std dev
man	all	0.468	0.155	0.438	0.157
	d	0.000	0.013	0.000	0.012
	$s=1$	0.451	0.157	0.421	0.159
	$s=2$	0.464	0.152	0.438	0.154
	$s=3$	0.482	0.156	0.452	0.158
	$s=4$	0.489	0.156	0.453	0.158
woman	all	0.312	0.095	0.321	0.094
	d	0.000	0.010	0.000	0.010
	$s=1$	0.357	0.092	0.371	0.091
	$s=2$	0.300	0.087	0.308	0.086
	$s=3$	0.299	0.090	0.311	0.088
	$s=4$	0.267	0.090	0.282	0.088
children	all	0.220	0.084	0.238	0.091
	d	0.000	0.006	0.000	0.006
	$s=1$	0.192	0.082	0.208	0.089
	$s=2$	0.236	0.083	0.251	0.089
	$s=3$	0.219	0.081	0.238	0.087
	$s=4$	0.211	0.082	0.265	0.088

for households with varying numbers of children. The pattern of children's shares rising at the expense of women's shares conditional on demographic characteristics (observed in Table 2) remains evident here where we average across those demographic characteristics.

5.4 Estimates of Random Resource Shares

Here we provide estimates of the distribution of the unobserved random variation in resources shares, based on the estimated models reported above. For each household member type j, given a value of s_j we numerically invert equation (10) to obtain U_{jh} for each household h. We then construct a new value for s_j as the sample average of $U_{jh} \ln U_{jh}$ across households, and iterate until these sample averages of $U_{jh} \ln U_{jh}$ converge. However, for households that have μ_{jh} close to zero, the resulting value of U_{jh} becomes very large, resulting in very large sample values of $U_{jh} \ln U_{jh}$ which then destabilizes the iteration procedure. To avoid this, we trim the data by dropping observations where $\mu_{jh} < 0.01$ for any household member (this entailed dropping just three and two households respectively, for our two models).

In Table 4, we present, separately for men, women and children, summary statistics of the distribution of values of U_{jh}, and the implied values of resource shares η_{jh}. We also give statistics for μ_{jh} for comparison. This table incorporates a number of checks on the model.

Our model assumes that U_{jh} are mean 1 and independent of p, x, z, d. However, as a check on the model we do not impose these features in the estimation or inversion. The sample average values of U_{jh} for men, women and children are 1.07, 1.05 and 0.84, respectively, which are reasonably close to their theoretical value of 1. Below the means, we report the correlations between the random resource share shifters, U_{jh}, and the deterministic part of resource shares, μ_{jh}. These correlations are 0.14, 0.03 and −0.06, respectively, for men, women and children, which are reasonably close to zero. Another implication of the model that we do not impose is that the mean of the Berkson errors e equals one. Our estimated sample mean of e is 1.02, which is very close to what the model predicts. Our estimated standard deviation of e is 0.73. Overall, these results suggest that strong distribution restrictions implied by our model appear to be satisfied.

Table 4: Random Resource Shares

		DF (2596 obs)		SAP (2597 obs)	
		mean	std dev	est	std dev
man	η_{man}	0.493	0.278	0.477	0.277
	μ_{man}	0.468	0.155	0.438	0.157
	d	0.000	0.013	0.000	0.012
	U_{man}	1.071	1.036	1.077	1.031
	$U_{man} \ln U_{man}$	0.494	1.643	0.494	1.637
	$\text{corr}(U_{man}, \mu_{man})$	0.139		0.140	
	$\text{corr}(\eta_{man}, \mu_{man})$	0.671		0.680	
woman	η_{woman}	0.321	0.228	0.320	0.221
	μ_{woman}	0.312	0.095	0.324	0.094
	d	0.000	0.010	0.000	0.010
	U_{woman}	1.055	1.025	1.022	0.993
	$U_{woman} \ln U_{woman}$	0.512	1.530	0.467	1.431
	$\text{corr}(U_{woman}, \mu_{woman})$	0.027		0.022	
	$\text{corr}(\eta_{woman}, \mu_{woman})$	0.571		0.530	
children	$\eta_{children}$	0.186	0.178	0.203	0.190
	$\mu_{children}$	0.220	0.084	0.238	0.091
	d	0.000	0.006	0.000	0.006
	$U_{children}$	0.836	0.904	0.832	0.885
	$U_{children} \ln U_{children}$	0.276	1.283	0.265	1.229
	$\text{corr}(U_{children}, \mu_{children})$	-0.059		-0.060	
	$\text{corr}(\eta_{children}, \mu_{children})$	0.396		0.401	

Now we turn to the distribution of resource shares η_{jh}. As we saw in Table 3, there is substantial variation in resource shares driven by observed covariates. In Table 4, we consider how much additional variation in resource shares emerges when we allow for possible unobserved distribution factors. Recall that $\eta_{jh} = \mu_{jh} U_{jh}/e_h$, where $\mu_{jh} = \mu_j(d_h, z_h)$ is the deterministic part of η_{jh} (varying by demographics and distribution factors), while U_{jh}/e_h is

the idiosyncratic random component of the resource shares. For men, the standard deviation of μ_{jh} is 0.155, while the standard deviation of η_{jh} is 0.277. So accounting for random component U_{jh}/c_h of resource shares (and hence accounting for unobserved distribution factors) roughly doubles the observed variation in resource shares. The estimates for women show a similar pattern. The standard deviation of μ_{jh} is 0.095 but that of η_{jh} is 0.228, roughly twice as large. For children, the standard deviation of μ_{jh} is 0.084 but that of η_{jh} is 0.177. These results suggest that using richer data sets to find additional determinants of intra-household inequality (i.e., additional distribution factors) may be an important avenue for future policy research.

These random resource share estimates were obtained using a model that was identified by Theorem 1, based on distribution factors. The right-hand columns of Table 4 gives comparable values for the resource share distribution in the model that imposes SAP, and so has levels identified by DLP. Here, we see much the same pattern. Random resource shares have roughly twice the standard deviation of resource shares that vary only with observed covariates. The bottom line is that unobserved distribution factors appear to be roughly as important as observed covariates in determining the variation in resource shares.

5.5 Estimated Poverty Rates

In our data, we observe that 95.3 per cent of households in our sample have household total expenditures of less than a household-level poverty line (defined as the sum of US$2 per adult plus US$1.20 per child) per day. Households vary by size, and larger households are more likely than smaller ones to fall below this threshold. As a result, 91.4 per cent of people in our sample live in households with household total expenditures below the threshold. However, this calculated povery rate ignores the fact that individuals within households may have unequal access to household resources. As noted in DLP, identification of the levels of resource shares allows us to calculate poverty rates at the person level, rather than at the household level, in a way that accounts for variation in access to household resources. In this subsection, we go beyond DLP to show the additional effect that accounting for random variation in resource shares has on the measurement of poverty at the individual level.

Table 5 gives estimated poverty rates for men, women and children using just the deterministic (fixed) part of resource shares, μ_{jh}, and also gives poverty rates that account for the random component of resource shares, η_{jh}. To estimate the poverty rate for men, for example, for each household h we compute the father's estimated resources, equal to $\mu_{man,h}$ times x_h or $\eta_{man,h}$ times x_h (the former for fixed resource shares and the latter for random resource shares), and we count the fraction of households for which these estimated men's

resources are less than the poverty threshold. For children, the personal budget of each child in a household is equal to $\mu_{children,h}$ times x_h/s or $\eta_{children,h}$ times x_h/s, where s is the number of children in the household. The poverty thresholds we use are US$2 per day for men and for women, and US$1.20 per day (60 per cent of the adult world poverty line, as implied by the Organisation for Economic Co-operation and Development (OECD) standard equivalence scales) for children.

Table 5: Measured Poverty

		DF (2596 obs)		SAP (2597 obs)	
		fixed	random	fixed	random
person level	men	58.7%	56.4%	62.2%	58.0%
	women	86.1%	79.0%	84.0%	79.0%
	children	97.6%	95.2%	96.9%	94.0%

Similar to DLP, we find much lower poverty rates for men than for women or for children. In Table 5 the estimated poverty rates based on fixed resource shares μ_{jh} are 59 per cent for men, 86 per cent for women and 98 per cent for children. Estimates that impose SAP are very similar.

Allowing for random heterogeneity in resource shares results in lower estimated poverty rates. Our estimated poverty rates based on η_{jh}, which include the random component U_{jh}, are 56 per cent for men, 79 per cent for women and 95 per cent for children. These estimated rates are 3 per cent lower for men and for children and 7 per cent lower for women. Very similar numbers are obtained from the model that additionally imposes SAP.

The reason allowing for random variation in resource shares tends to lower estimated poverty rates is because the poverty line is above the mode of the (unimodal) μ_{jh} distribution for each j. The random component U_{jh} moves some fraction of people who had μ_{jh} near to but above the poverty line down below it, and pushes a different fraction of people who were below the line up above it. The poverty line is in the downward sloping part of the density of μ_{jh}, so there are more people close to but below the line than there are people close to but above the line, and hence more who are randomly moved from below to above than those who are randomly moved from above to below.

The reduction in estimated poverty from allowing for the random component of resource shares is larger for women than for men or children for two reasons. First, the slope of the estimated density in the neighborhood of the poverty line is steeper for women. And second, the random component of resource shares is smaller for children. Looking back at Table 4, allowing for random resource shares increases the standard deviation of women's resource shares by 0.13, but only increases the standard deviation of children's resource shares by 0.09.

We can summarize these results as follows. Calculating poverty rates in the standard way, at the household level, essentially assumes equal sharing within the household and thereby ignores actual intra-household allocation. Ignoring intra-household allocations greatly overestimates men's poverty rates and greatly underestimates the poverty rates of women and children, as reported by DLP and others (and reconfirmed here). However, failing to account for random variation in resource shares results in estimated poverty rates that are biased upwards. Essentially, some individuals who seem poor on the basis of their observed characteristics (demographics and distribution factors) are in fact non-poor due to unobserved characteristics.

6 Conclusions

In this paper, we consider random resource shares and a new way of identifying the level of resource shares. In Theorem 1 we show identification of the distribution (around an unknown level) of resource shares that are permitted to vary randomly across households. In Theorem 2 we show identification of the level of resource shares that does not require previously needed assumptions regarding similarity of tastes across households of varying compositions or across people within a household. This second theorem uses a new way to exploit variation in so-called distribution factors, that is, variables that affect the allocation of resources within households but do not affect the tastes of individuals.

We apply our results to household survey data from Malawi. We find that identification using distribution factors is empirically practical, but can result in imprecise estimates if distribution factors are too highly correlated with other regressors. We use our results to test some of the behavioural restrictions that previous results required for identification. We find that the SAP restriction proposed by Dunbar, Lewbel and Pendakur (2013) for identification is an acceptable restriction and can be combined with observed distribution factors to yield more efficient (and numerically better behaved) estimates.

Our estimates of the distribution of random resource shares show that unobserved factors account for about as much variation in resource shares as do observed covariates. We find that accounting for unobserved factors (i.e., random variation) in resource shares across household members matters a great deal for poverty analyses. In particular, accounting for such unobserved heterogeneity results in lower estimated poverty rates.

References

[1] Banks, J., R. Blundell and A. Lewbel, 1997. "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79(4), 527-539.

[2] Bargain, O. and O. Donni, 2009. "The Measurement of Child Costs: A Rothbarth-Type Method Consistent with Scale Economies," Institute for the Study of Labor (IZA) Discussion Paper No. 4654.

[3] Bargain, O. & O. Donni, 2012. "Expenditure on children: A Rothbarth-type method consistent with scale economies and parents' bargaining," *European Economic Review*, 56(4), 792-813.

[4] Becker, G. S., 1965. "A Theory of the Allocation of Time," *The Economic Journal*, 75(299), 493-517.

[5] Becker, G. S., 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

[6] Blundell, R., P.-A. Chiappori and C. Meghir, 2005. "Collective Labor Supply with Children," *Journal of Political Economy*, 113(6), 1277-1306.

[7] Bourguignon, F. and P.-A. Chiappori, 1992. "Collective models of household behavior : An introduction," *European Economic Review*, 36(2-3), 355-364.

[8] Browning, M., F. Bourguignon and P.-A. Chiappori and V. Lechene, 1994. "Income and Outcomes: A Structural Model of Intrahousehold Allocation," *Journal of Political Economy*, 102(6), 1067-1096.

[9] Browning, M. and P.-A. Chiappori, 1998. "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests," *Econometrica*, 66(6), 1241-1278.

[10] Browning, M., P.-A. Chiappori and A. Lewbel, 2013. "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, 80(4), 1267-1303.

[11] Calvi, R., 2017. "Why Are Older Women Missing in India? The Age Profile of Bargaining Power and Poverty," mimeo.

[12] Cherchye, L., B. De Rock and F. Vermeulen, 2011. "The revealed preference approach to collective consumption behavior: testing and sharing rule recovery," *Review of Economic Studies*, 78(1), 176-198.

[13] Cherchye, L., B. De Rock and F. Vermeulen, 2012. "Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information," *American Economic Review*, 102(7), 3377-3405.

[14] Cherchye, L., B. De Rock, T. Demuynck, and F. Vermeulen, 2014. "Household consumption when the marriage is stable," Unpublished manuscript.

[15] Cherchye, L., B. De Rock, A. Lewbel and F. Vermeulen, 2015. "Sharing Rule Identification for General Collective Consumption Models," *Econometrica*, 83(5), 2001-2041.

[16] Chiappori, P.-A., 1988. "Nash-Bargained Households Decisions: A Comment," *International Economic Review*, 29: 791-796.

[17] Chiappori, P.-A., 1992. "Collective Labor Supply and Welfare," *Journal of Political Economy*, 100(3), 437-467.

[18] Chiappori, P.-A., B. Fortin and G. Lacroix, 2002. "Marriage Market, Divorce Legislation, and Household Labor Supply," *Journal of Political Economy*, 110(1), 37-72.

[19] Chiappori, P.-A. and I. Ekeland, 2009. "The Microeconomics of Efficient Group Behavior: Identification," *Econometrica*, 77(3), 763-799.

[20] Chiappori, P.-A. and J. H. Kim, 2017. "Identifying Heterogeneous Sharing Rules," *Quantitative Economics*, 8(1), 201-218.

[21] Couprie, H., E. Peluso and A. Tranmoy, 2010. "Is power more evenly balanced in poor households?" *Journal of Public Economics*, 94(7-8), 493-507.

[22] Deaton, A. and J. Muellbauer, 1980. "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312-326.

[23] Dunbar, G., A. Lewbel and K. Pendakur, 2013. "Children's Resources in Collective Households: Identification, Estimation and an Application to Child Poverty in Malawi," *American Economic Review*, 103, 438-471.

[24] Jorgenson, D. W., L. J. Lau and T. M. Stoker (1982). "Transcendental Logarithmic Model of Aggregate Consumer Behavior," in *Aggregate Consumer Behavior*, Cambridge, The MIT Press, 203-356.

[25] Lewbel, A., 2003. "A rational rank four demand system," *Journal of Applied Econometrics*, 18(2), 127-135.

[26] Lewbel, A. and K. Pendakur, 2008. "Estimation of Collective Household Models With Engel Curves," *Journal of Econometrics*, 147, 350-358.

[27] Lise, J. and S. Seitz, 2011. "Consumption Inequality and Intra-Household Allocations," *Review of Economic Studies*, 2011(1), 328-355.

[28] Matzkin, R. and J. Perez-Estrada, 2011. "Heterogeneous Collective Demand: Exploring Differential Generosity within Elderly Couples," mimeo.

[29] Menon, M., K. Pendakur and F. Perali, 2012. "On the Expenditure-Dependence of Resource Shares," *Economics Letters*, 117, 739-742.

[30] Muellbauer, J. N. J., 1976. "Community Preference Fields and the Representative Consumer," *Econometrica*, 44(5), 979-999.

[31] Vermeulen, F., 2002. "Collective Household Models: Principles and Main Results," *Journal of Economic Surveys*, 16(4), 533-564.

A Appendix: Proofs and Additional Results

This Appendix has three sections. The first is derivation of the household demand equations for the private assignable goods. The second gives proofs of our theorems. The third describes an estimator of the resource share distribution for general demand equation functional forms, instead of the estimator given in the text that was specific to the class of functional forms we use for our empirical analysis.

A.1 Derivation of household demand equations of private assignable goods

Here we summarize the derivation of equation (1), which is based on the same machinery used in BCL and DLP to analyze collective household models where most goods can be shared to varying extents. Unlike DLP, we explicitly include price variation; unlike BCL, we focus on assignable goods and include more than two household members; and unlike both DLP and BCL, we explicitly include distribution factors, both observable d and unobservable v.

Start by assuming the household chooses what to consume using the program

$$\max_{y_1,...,y_J:r_1,...,r_J} \widetilde{V}\left[V_1\left(y_1,r_1,z\right),...,V_J\left(y_J,r_J,z\right) \mid d,z,p/x^*,v\right] \tag{11}$$

$$\text{such that } r = \sum_{j=1}^{J} r_j \text{ and } x^e = p_r' A(z) r + \sum_{j=1}^{J} p_j y_j$$

where $V_j(y_j, r_j, z)$ for $j = 1, \ldots, J$ is the utility function of household member j, and the function $\tilde{V}$ describes the social welfare function or bargaining process of the household, which exists because the household is Pareto efficient.

Recall z denotes a vector of observable attributes of households and their members like age, education and number of children. Household attributes z may affect preferences, and so appear inside the utility functions V_j. These z variables may also affect the bargaining process or social welfare function given by $\tilde{V}$ (by, e.g., affecting the relative bargaining power of members), and as a result may affect resource shares.

We have scalars $y_1, \ldots, y_J$ that are the quantities of private assignable goods, where member j has quantity y_j in his or her utility function, and does not have y_ℓ for all $\ell \neq j$ in his or her utility function. Each member's utility function also depends on a quantity vector of other goods r_j. The market prices of these goods are given by the vector p_r. The square matrix $A(z)$ is what BCL call a linear consumption technology function over goods. Having $A(z)$ differ from the identity matrix is what allows goods in r to be partly shared and/or consumed jointly. In particular, $A(z)r$ equals the quantity vector of these goods that the household actually purchases, while $r = \sum_{j=1}^{J} r_j$ is total quantity vector of these goods that the household consumes. These quantities are not the same due to sharing and joint consumption. The smaller an element of $A(z)r$ is relative to the corresponding element of r, the more that good is shared or jointly consumed. See BCL for details. The vector of all prices p includes p_r, the vector of prices of the elements of r, and $p_1, \ldots, p_J$, the prices of the private assignable goods $y_1, \ldots, y_J$.

What makes the vectors d and v be distribution factors (observed and unobserved, respectively) in the model is that they appear only as arguments of $\tilde{V}$, and so affect only the allocation of resources within the household, but not the tastes of the individual household members.

Applying duality theory and decentralization welfare theorems, it follows from BCL that the household's program above is equivalent to a program where each household member j chooses what to consume using the program

$$\max_{y_j, r_j} V_j(y_j, r_j, z) \text{ such that } \eta_j(p, x^e, z, d, v) x^e = p_r' A(z) r_j + p_j y_j \tag{12}$$

where $\eta_j = \eta_j(p, x^e, z, d, v)$ is the resource share of member j; that is, η_j is the fraction of total household resources x^e that are allocated to member j. This member then chooses quantities y_j and the vector r_j subject to a linear budget constraint. The vector $p_r A(z)$

equals the vector of shadow prices of goods r. These shadow prices for the household are lower than market prices, due to sharing. Being private and assignable, the shadow price of each y_j equals its market price p_j. The shadow budget for member j is $\widetilde{x}_j = \eta_j x^*$. As shown in BCL, the resource share functions $\eta_j(p, x^*, z, d, v)$ for each member j in general depend on the function $\widetilde{V}$ and on the utility functions $V_1,...,V_J$.

BCL show that the more bargaining power a household member has (i.e., the greater is the weight of his or her utility function in $\widetilde{V}$), the larger is their resource share η_j. Resource shares η_j all lie between zero and one, and resource shares sum to one, that is, $\sum_{j=1}^{J} \eta_j = 1$.

As in DLP, we will not work with the household demand functions of all goods (which, as shown in BCL, can be rather complicated). Instead, we only make use of the demand functions of the private assignable goods y_j, which are simpler. Since equation (12) is an ordinary utility function maximized under a linear budget constraint (linear in shadow prices and a shadow budget), the solution to equation (12) is a set of Marshallian demand equations for y_j and r_j. Let $h_j(\widetilde{x}_j, p, z)$ be the Marshallian demand function of person j' for their private assignable good, that is, $h_j(\widetilde{x}_j, p, z)$ is the quantity person j in a household with member attributes z would demand of their assignable good if they had a budget equal to their shadow budget $\widetilde{x}_j$ and faced the within-household shadow price vector that corresponds to the market price vector p. Since each y_j is private and assignable, the quantity y_j that member j chooses to consume equals the quantity of this good that the household buys. It therefore follows from the above that the household's quantity demand of each private assignable good y_j is given by equation (1).

A.2 Proofs

PROOF OF THEOREM 1: Let $h_j^{-1}(y_j, p, z)$ denote the inverse of the function h_j with respect to its first argument, which exists by Assumption A1. Let $F_{Y_j}(y_j \mid p, x, d, z)$ be the identified conditional distribution of Y_j. Let $\widetilde{\eta}_j = e\eta_j$ and let $F_{\widetilde{\eta}_j}(\widetilde{\eta}_j \mid p, d, z)$ be the unknown conditional distribution of $\widetilde{\eta}_j$. Then for any $y_j \in supp(Y_j \mid p, x, d, z)$,

$$F_{Y_j}(y_j \mid p, x, d, z) = \Pr(h_j(\widetilde{\eta}_j x, p, z) \leq y_j \mid p, x, d, z) = \Pr\left(\widetilde{\eta}_j \leq \frac{h_j^{-1}(y_j, p, z)}{x} \mid p, x, d, z\right)$$
$$= F_{\widetilde{\eta}_j}\left(\frac{h_j^{-1}(y_j, p, z)}{x} \mid p, x, d, z\right) = F_{\widetilde{\eta}_j}\left(\frac{h_j^{-1}(y_j, p, z)}{x} \mid p, d, z\right)$$

where the last equality follows from Assumption A3. By continuity of the distribution of $\widetilde{\eta}_j$, the distribution function $F_{\widetilde{\eta}_j}$ is differentiable, and

$$\frac{-x\partial F_{Y_j}\left(y_j \mid p,x,d,z\right)/\partial y_j}{\partial F_{Y_j}\left(y_j \mid p,x,d,z\right)/\partial x} = \frac{-x\partial h_j^{-1}\left(y_j,p,z\right)/\partial y_j}{-xh_j^{-1}\left(y_j,p,z\right)} = \frac{\partial \ln h_j^{-1}\left(y_j,p,z\right)}{\partial y_j}.$$

It follows by the continuity assumptions and interval support of x that $supp\left(Y_j \mid p,z\right)$ is an interval. Let $\kappa_j\left(p,z\right)$ denote any given nonzero element in this support. Define the identified function r_j by

$$\begin{aligned} r_j\left(y_j,p,z\right) &= \exp\left(\int_{\kappa_j(p,z)}^{y_j} E\left(\frac{-x\partial F_{Y_j}\left(y_j \mid p,x\right)/\partial y_j}{\partial F_{Y_j}\left(y_j \mid p,x\right)/\partial x} \mid y_j,p,z\right) dy_j\right) \\ &= \exp\left(\int_{\kappa_j(p,z)}^{y_j} \frac{\partial \ln h_j^{-1}\left(y_j,p,z\right)}{\partial y_j} dy_j\right) \\ &= h_j^{-1}\left(y_j,p,z\right) c_j\left(p,z\right) \end{aligned}$$

where $c_j\left(p,z\right) = 1/h_j^{-1}\left(\kappa_j\left(p,z\right),p,z\right)$ is an unknown function.

Now $h_j^{-1}\left(Y_j,p,z\right) = \widetilde{\eta}_j x$ so $r_j\left(Y_j,p,z\right)/x = \widetilde{\eta}_j c_j\left(p,z\right)$. Applying the same derivation used above to relate F_{Y_j} to $F_{\widetilde{\eta}_j}$, we have

$$\begin{aligned} F_{Y_1,...Y_J}\left(y_1,...y_J \mid p,x,d,z\right) &= F_{\widetilde{\eta}_1,...\widetilde{\eta}_J}\left(\frac{h_1^{-1}\left(y_1,p,z\right)}{x},...\frac{h_J^{-1}\left(y_J,p,z\right)}{x} \mid p,d,z\right) \\ &= F_{\widetilde{\eta}_1,...\widetilde{\eta}_J}\left(\frac{r_1\left(y_1,p,z\right)}{c_1\left(p,z\right)x},...\frac{r_J\left(y_J,p,z\right)}{c_J\left(p,z\right)x} \mid p,d,z\right). \end{aligned}$$

Since the $F_{Y_1,...Y_J}$ is known and the $r_j\left(Y_j,p,z\right)$ functions are identified, this shows that the $F_{\widetilde{\eta}_1,...\widetilde{\eta}_J}$ distribution is identified up to the unknown location terms $c_1\left(p,z\right),...,\ c_J\left(p,z\right)$. Equivalently, we can say that the joint distribution of $r_1\left(Y_j,p,z\right)/x,...\ r_J\left(Y_j,p,z\right)/x$ conditional on p,d,z is identified from identification of $F_Y\left(Y_1,...Y_j \mid p,x,d,z\right)$ and of the functions $r_j\left(y_j,p,z\right)$. This joint distribution of $r_1\left(Y_j,p,z\right)/x,...\ r_J\left(Y_j,p,z\right)/x$ conditional on p,d,z equals the joint distribution of $\widetilde{\eta}_1 c_1\left(p,z\right),...,\widetilde{\eta}_J c_J\left(p,z\right)$ conditional on p,d,z, which proves the theorem.

PROOF OF THEOREM 2: Define the identified function t_j by

$$\begin{aligned} t_j\left(p,d,z\right) &= E\left(\frac{r_j\left(Y_j,p,z\right)}{x} \mid p,d,z\right) = E\left(e\eta_j c_j\left(p,z\right) \mid p,d,z\right) \\ &= E\left(e\right) E\left(\eta_j \mid p,d,z\right) c_j\left(p,z\right) \end{aligned}$$

so

$$\frac{t_j(p,d_k,z)}{t_j(p,d_1,z)}E(\eta_j \mid p,d_1,z) = E(\eta_j \mid p,d_k,z).$$

It follows that $T(p,z)$ defined in Assumption A4 equals the J by J matrix that has $t_j(p,d_k,z)/t_j(p,d_1,z)$ in the row k and column j position, and therefore that the matrix $T(p,z)$ is identified. We also have

$$\sum_{j=1}^{J}\frac{t_j(p,d_k,z)}{t_j(p,d_1,z)}E(\eta_j \mid p,d_1,z) = \sum_{j=1}^{J}E(\eta_j \mid p,d_k,z) = E\left(\sum_{j=1}^{J}\eta_j \mid p,d_k,z\right) = 1$$

Let $E(\eta \mid p,d_1,z)$ be the vector of elements $E(\eta_j \mid p,d_1,z)$ for $j = 1,....,J$, and let 1_J denote the J vector of ones. Then the above equation for $k = 1,....,J$ is equivalent to $T(p,z)E(\eta \mid p,d_1,z) = 1_J$, and so, using Assumption A4, $E(\eta \mid p,d_1,z) = T(p,z)^{-1}1_J$, and therefore $E(\eta_j \mid p,d_1,z)$ is identified for $j = 1,....,J$. We can then identify the functions $c_j(p,z)$ for $j = 1,....,J$ by $c_j(p,z) = t_j(p,d_1,z)/E(\eta_j \mid p,d_1,z)$. By Theorem 1, the joint distribution of $\widetilde{\eta}_1 c_1(p,z),....,\widetilde{\eta}_J c_J(p,z)$ conditional on p,d,z, is identified, and since $c_j(p,z)$ for $j = 1,....,J$ is identified, we can divide elements of one by the other to conclude that the joint distribution of $\widetilde{\eta}_1,....\widetilde{\eta}_J$ conditional on p,d,z, is identified. Now $e = \widetilde{\eta}_1 + ,...., + \widetilde{\eta}_J$ and therefore the distribution of e is identified. We have that the joint distribution of $\ln\widetilde{\eta} = \ln\eta + \ln e$ conditional on p,d,z, is identified, and the distribution of $\ln e$ is identified and independent of $\ln\eta$, so by a deconvolution the joint distribution of $\ln\eta$ and hence of η, conditional on p,d,z, is identified.

A.3 A General Estimator

In the text we provide an estimator for the model that depends in part on features of our chosen piglog functional form for Engel curves. Here we describe a more general estimation method that doesn't depend on specific properties of the piglog but is more complicated to apply in that it requires estimating an inverted demand system. Suppress z for the moment: both h_j and η_j depend on it. Suppress p also and for now consider an Engel curve exercise. Assume we have some parameterization of the clothing demand functions h_j, so the model is

$$Y_j = h_j(x\eta_j e, z \mid \theta_j),$$

where θ_j is a vector of clothing engel curve parameters, η_j are random resource shares, and e is multiplicative measurement error on total expenditures.

For $j = 1,2,3$ assume that $E(\eta_j \mid x,d,z) = E(\eta_j \mid d,z)$, so on average resource shares

do not depend on x. Let

$$s_j(d, z \mid \gamma) = E(\eta_j \mid d, z),$$

so s_j is an assumed parameterized functional form for mean resource shares, which depends on a parameter vector γ. Resource shares sum to one, that is, $\eta_1 + \eta_2 + \eta_3 = 1$, so $s_1(d, z \mid \gamma_1) + s_2(d, z \mid \gamma_2) + s_3(d, z \mid \gamma_3) = 1$.

We assume e is a multiplicative measurement error, with $E[e \mid d, x, z, \eta_1, \eta_2, \eta_3] = 1$. Let Y_j be normal goods, so that h_j is invertible. Let $g_j(.) = h_j^{-1}(.)$ be the inverse demand functions for each j. Then by definition

$$0 = g_j(Y_j \mid \theta_j) - x\eta_j e.$$

Taking conditional expectations gives

$$\begin{aligned} 0 &= E[g_j(Y_j \mid \theta_j) - x\eta_j e \mid d, x, z, \eta_1, \eta_2, \eta_3] \\ &= E[g_j(Y_j \mid \theta_j) \mid d, x, z, \eta_1, \eta_2, \eta_3] - x\eta_j E[e \mid d, x, z, \eta_1, \eta_2, \eta_3] \\ &= E[g_j(Y_j \mid \theta_j) \mid d, x, z, \eta_1, \eta_2, \eta_3] - x\eta_j. \end{aligned}$$

Now take conditional expectations of this, just conditioning on d, x, z, and apply the law of iterated expectations to get

$$\begin{aligned} 0 &= E[g_j(Y_j \mid \theta_j) \mid d, x, z] - xE[\eta_j \mid d, x, z] \\ &= E[g_j(Y_j \mid \theta_j) \mid d, x, z] - xE[\eta_j \mid d, z] \\ &= E[g_j(Y_j \mid \theta_j) \mid d, x, z] - xs_j(d, z \mid \gamma) \\ &= E[g_j(Y_j \mid \theta_j) - xs_j(d, z \mid \gamma) \mid d, x, z]. \end{aligned}$$

Based on this last equation estimate, the parameters θ_j and γ for $j = 1, 2, 3$ by GMM, using unconditional moments of the form

$$0 = E[[g_j(Y_j \mid \theta_j) - xs_j(d, z \mid \gamma)]\, r(d, x, z)] = 0$$

for $j = 1, 2, 3$, where instruments $r(x, d, z)$ are functions of the variables x, d and z that we choose, and g_j and s_j have the functional forms we choose.

Now

$$g_1(Y_1 \mid \theta_j) + g_2(Y_2 \mid \theta_2) + g_3(Y_3 \mid \theta_3) = (\eta_1 + \eta_2 + \eta_3)\, ex = ex.$$

So, given estimates of the parameters $\widehat{\theta}_j$, we can recover estimates of the value of e, η_1, η_2,

and η_3 for each consumer j in each household i by

$$\widehat{e}_i = \frac{g_1(Y_{1i} \mid \widehat{\theta}_j) + g_2\left(Y_{2i} \mid \widehat{\theta}_2\right) + g_3(Y_{3i} \mid \widehat{\theta}_3)}{x_i}$$

and

$$\widehat{\eta}_{ji} = \frac{g_j(Y_{ji} \mid \theta_j)}{\widehat{e}_i x_i}.$$

We could estimate the population distribution of e and the joint distribution of resource shares by kernel density estimation using the above estimates of each consumer's $\widehat{e}_i$, and $\widehat{\eta}_{ji}$ as data. Also, given $\widehat{e}_i$, and $\widehat{\eta}_{ji}$, the demand function of a consumer j in household i is given by $h_j(\widehat{e}_i \widehat{\eta}_{ji} x_i \mid \widehat{\theta}_j)$.

A.4 Additional Results

Appendix Table A1: Detailed Parameter Estimates, man

		BITE		SAP	
		est	std err	est	std err
α_{man}	constant	0.347	0.058	0.316	0.054
	$s = 2$	0.009	0.031	0.016	0.026
	$s = 3$	0.028	0.032	0.031	0.028
	$s = 4$	0.015	0.044	0.014	0.034
	North	0.369	0.054	0.378	0.055
	Central	0.233	0.039	0.235	0.036
	avg (child age) -5	0.037	0.045	0.031	0.041
	min (child age) -5	-0.032	0.027	-0.030	0.026
	man age -28	0.046	0.104	0.037	0.097
	woman age -22	-0.131	0.116	-0.136	0.108
	man education	0.051	0.063	0.056	0.062
	woman education	0.004	0.091	0.012	0.089
	ln(dist to road)	0.134	0.034	0.134	0.035
	ln(dist daily mkt)	-0.170	0.054	-0.177	0.050
	dry season	-0.133	0.025	-0.130	0.024
	Christian	0.081	0.042	0.081	0.038
	Muslim	0.111	0.050	0.114	0.047
	d_1 first princ. comp.	0.007	0.005	0.007	0.004
	d_2 second princ. comp.	-0.007	0.004	-0.007	0.004
	d_3 third princ. comp.	0.005	0.005	0.004	0.005
γ_{man}	constant	0.059	0.009	0.067	0.009
	$s = 2$	-0.002	0.005	-0.002	0.005
	$s = 3$	-0.005	0.005	-0.006	0.005
	$s = 4$	-0.010	0.005	-0.012	0.005
	North	-0.026	0.005	-0.032	0.006
	Central	-0.024	0.005	-0.029	0.005
	avg (child age) -5	-0.004	0.007	-0.003	0.007
	min (child age) -5	0.008	0.004	0.009	0.004
	man age -28	-0.027	0.011	-0.028	0.012
	woman age -22	0.011	0.012	0.012	0.013
	man education	-0.013	0.008	-0.015	0.008
	woman education	0.003	0.011	0.001	0.011
	ln(dist to road)	-0.017	0.004	-0.019	0.005
	ln(dist daily mkt)	0.010	0.006	0.012	0.006
	dry season	0.026	0.005	0.029	0.004
	Christian	-0.012	0.006	-0.013	0.007
	Muslim	-0.015	0.007	-0.017	0.008
β_{man}	constant	0.016	0.003	0.018	0.003
	$s = 2$	0.005	0.003	0.005	0.004
	$s = 3$	0.006	0.004	0.007	0.004
	$s = 4$	0.002	0.004	0.000	0.004

Appendix Table A1: Detailed Parameter Estimates, woman

		BITE		SAP	
		est	std err	est	std err
α_{woman}	constant	0.392	0.098	0.396	0.049
	$s = 2$	-0.065	0.021	-0.074	0.022
	$s = 3$	-0.072	0.023	-0.080	0.023
	$s = 4$	-0.095	0.027	-0.102	0.027
	North	-0.195	0.071	-0.182	0.043
	Central	-0.088	0.042	-0.076	0.028
	avg (child age) -5	0.022	0.035	0.031	0.035
	min (child age) -5	0.011	0.022	0.011	0.021
	man age -28	-0.012	0.103	-0.012	0.095
	woman age -22	0.038	0.108	0.051	0.101
	man education	-0.002	0.050	-0.001	0.051
	woman education	-0.071	0.068	-0.087	0.070
	ln(dist to road)	-0.083	0.029	-0.085	0.027
	ln(dist daily mkt)	0.172	0.047	0.179	0.042
	dry season	0.101	0.027	0.099	0.020
	Christian	-0.069	0.037	-0.068	0.033
	Muslim	-0.098	0.044	-0.094	0.041
	d_1 first princ. comp.	-0.007	0.004	-0.007	0.004
	d_2 second princ. comp.	0.002	0.003	0.002	0.003
	d_3 third princ. comp.	-0.002	0.004	-0.001	0.004
γ_{woman}	constant	0.077	0.031	0.072	0.015
	$s = 2$	0.008	0.012	0.016	0.007
	$s = 3$	0.010	0.013	0.011	0.007
	$s = 4$	-0.009	0.013	0.004	0.007
	North	0.079	0.027	0.069	0.023
	Central	0.027	0.009	0.023	0.007
	avg (child age) -5	-0.008	0.009	-0.009	0.008
	min (child age) -5	-0.002	0.006	-0.002	0.006
	man age -28	0.009	0.021	0.010	0.020
	woman age -22	-0.052	0.025	-0.052	0.022
	man education	0.000	0.017	0.000	0.017
	woman education	0.016	0.022	0.019	0.021
	ln(dist to road)	0.012	0.009	0.012	0.009
	ln(dist daily mkt)	-0.051	0.020	-0.050	0.014
	dry season	-0.004	0.005	-0.002	0.005
	Christian	0.010	0.007	0.009	0.007
	Muslim	0.030	0.013	0.028	0.010
β_{woman}	constant	0.023	0.009		
	$s = 2$	-0.001	0.008		
	$s = 3$	0.006	0.008		
	$s = 4$	-0.011	0.010		

Appendix Table A1: Detailed Parameter Estimates, children

		BITE		SAP	
		est	std err	est	std err
$\gamma_{children}$	constant	0.049	0.024	0.052	0.013
	$s = 2$	0.064	0.033	0.033	0.009
	$s = 3$	0.063	0.033	0.048	0.009
	$s = 4$	0.052	0.032	0.032	0.010
	North	0.122	0.061	0.120	0.045
	Central	0.048	0.025	0.046	0.015
	avg (child age) -5	0.009	0.010	0.008	0.009
	min (child age) -5	0.002	0.006	0.003	0.005
	man age -28	0.010	0.021	0.007	0.019
	woman age -22	-0.052	0.029	-0.043	0.021
	man education	0.044	0.022	0.043	0.016
	woman education	-0.003	0.022	-0.004	0.019
	ln(dist to road)	0.022	0.012	0.017	0.009
	ln(dist daily mkt)	-0.013	0.013	-0.012	0.011
	dry season	0.007	0.006	0.008	0.004
	Christian	0.006	0.007	0.006	0.006
	Muslim	0.011	0.009	0.012	0.008
$\beta_{children}$	constant	0.013	0.007		
	$s = 2$	0.021	0.011		
	$s = 3$	0.014	0.010		
	$s = 4$	0.009	0.011		